IEEE JOURNAL OF

SOLID-STATE CIRCUITS

A PUBLICATION OF THE IEEE SOLID-STATE CIRCUITS SOCIETY

FEBRUARY 2005 VOLUME 40 NUMBER 2 IJSCBC (ISSN 0018-9200)

(Contents Continued on Back Cover)

IEEE

IEEE JOURNAL OF SOLID-STATE CIRCUITS

The IEEE JOURNAL OF SOLID-STATE CIRCUITS is published by the IEEE Solid-State Circuits Society. All IEEE members are eligible for membership and will receive this JOURNAL upon payment of the annual society membership fee of $18.00. For information on receiving this JOURNAL, write to the IEEE Service Center at the address below. *Member copies of Transactions/Journals are for personal use only.*

IEEE JOURNAL OF SOLID-STATE CIRCUITS (ISSN 0018-9200) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society/Council, or its members. **IEEE Corporate Office:** 3 Park Avenue, 17th Floor, New York, NY 10016-5997. **IEEE Operations Center:** 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. **NJ Telephone:** +1 732 981 0060. **Price/Publication Information:** Individual copies: IEEE members $20.00 (first copy only), nonmembers $48.00 per copy. (Note: Postage and handling charge not included.) Member and nonmember subscription prices available on request. Available on CD-ROM and DVD (see http://www.sscs.org/jssc/) as well as in microfiche and microfilm.

Periodicals Postage Paid at New York, NY, and at additional mailing offices. **Postmaster:** Send address changes to IEEE JOURNAL OF SOLID-STATE CIRCUITS, IEEE, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. GST Registration No. 125634188. Printed in U.S.A.

Digital Object Identifier 10.1109/JSSC.2005.844051

IEEE JOURNAL OF

SOLID-STATE CIRCUITS

A PUBLICATION OF THE IEEE SOLID-STATE CIRCUITS SOCIETY

FEBRUARY 2005 VOLUME 40 NUMBER 2 IJSCBC (ISSN 0018-9200)

IEEE

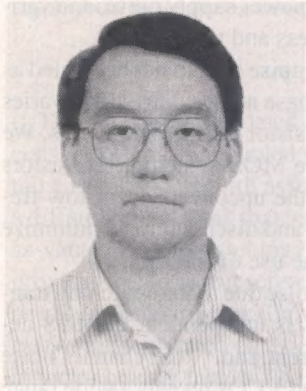
New Associate Editor

IT is my pleasure to announce the appointment of Dr. Jerry (Heng-Chih) Lin as an Associate Editor. Dr. Lin brings with him a vast amount of experience in analog, wireless and wireline communication circuits, all of which continue to be areas of active research and development. He will certainly be an asset to the JOURNAL.

Dr. Michael Perrott has retired. I would like to thank Dr. Perrott for his outstanding service to the JOURNAL. He will be missed.

KRISHNASWAMY NAGARAJ, *Editor-in-Chief*
Texas Instruments, Inc.
Dallas, TX 75243 USA

Digital Object Identifier 10.1109/JSSC.2004.842371



Jerry (Heng-Chih) Lin (M'89–SM'03) received the B.S. and M.S. degrees from National Taiwan University, Taipei, Taiwan, and the Ph.D. degree from Stanford University, Stanford, CA, in 1989, 1991, and 1997, respectively, all in electrical engineering.

From 1991 to 1993, he served as a 2nd rank Lieutenant in the Taiwanese army. He joined Texas Instruments, Inc., Dallas, TX, in 1997 and is currently a Member, Group Technical Staff. His research interests include wireless and wireline transceivers, phase-locked loops, low-power low-voltage ADCs and DACs in nanometer-scaled CMOS processes. He has more than 20 publications and has more than 20 U.S. patents granted or pending.

Dr. Lin chaired the Dallas Chapter of the IEEE Solid-State Circuits Society in 2002.

The Impact of Device Type and Sizing on Phase Noise Mechanisms

Albert Jerng, *Student Member, IEEE,* and Charles G. Sodini, *Fellow, IEEE*

***Abstract*—Phase noise mechanisms in integrated LC voltage-controlled oscillators (VCOs) using MOS transistors are investigated. The degradation in phase noise due to low-frequency bias noise is shown to be a function of AM-PM conversion in the MOS switching transistors. By exploiting this dependence, bias noise contributions to phase noise are minimized through MOS device sizing rather than through filtering. NMOS and PMOS VCO designs are compared in terms of thermal noise. Short-channel MOS considerations explain why 0.18-μm PMOS devices can attain better phase noise than 0.18-μm NMOS devices in the $1/f^2$ region. Phase noise in the $1/f^3$ region is primarily dependent upon the upconversion of flicker noise from the MOS switching transistors rather than from the bias circuit, and can be improved by decreasing MOS switching device size. Measured results on an experimental set of VCOs confirm the dependencies predicted by analysis. A 5.3-GHz all-PMOS VCO topology demonstrates measured phase noise of -124 dBc/Hz at 1-MHz offset and -100 dBc/Hz at 100-kHz offset while dissipating 13.5 mW from a 1.8-V supply using a 0.18-μm SiGe BiCMOS process.**

***Index Terms*—Flicker noise, phase noise, voltage-controlled oscillator (VCO), WiGLAN.**

I. Introduction

HIGH data-rate wireless LAN applications are driving the continued development of highly integrated system-on-chip (SOC) solutions. Integrated voltage-controlled oscillators (VCOs) are essential components of such wireless systems. In this work, the VCOs are designed in the context of the MIT Wireless Gigabit Local Area Network (WiGLAN) project. The aim of the WiGLAN is to achieve a maximum 1-Gb/s data rate using 150 MHz of bandwidth in frequency bands allocated in the 5–6-GHz range. An adaptive M-ary modulation scheme, up to 256 QAM, is chosen to achieve higher data rates, imposing stringent accuracy requirements on the local oscillator (LO) signal. Thus, VCOs with low phase noise and high operating frequencies are required.

The nonlinear and time-varying nature of an oscillator complicates phase noise analysis [1]. The existence of many sources of noise coming from several different frequencies makes it difficult to discern which noise mechanisms are the dominant ones. Recent progress in VCO research has revealed that bias noise is an important contributor to phase noise [2]. High-frequency bias current noise has been observed to be a dominant contributor to phase noise [3], [4]. AM noise originating from the upconversion of low-frequency bias noise cannot be neglected due to AM-PM conversion through the varactor [5]–[7]. Solutions to reduce bias noise have included filtering [4], [5], and removing the bias current generator [6]. Filtering requires extra inductors and capacitors. Removing the current source is also problematic, due to increased sensitivity to power supply noise, and variation of the bias current over process and temperature.

In Section III, we show that bias noise should not be treated as a fixed source of phase noise. Its phase noise contribution varies as a function of the switching transistor device parameters. We identify AM-PM conversion in the MOS switching transistors as a fundamental mechanism for the upconversion of low-frequency bias noise into phase noise and discuss how to minimize the upconversion factor without the use of filtering.

It has been shown that thermal noise due to the switching transistors in CMOS implementations is independent of MOS device size and depends on bias current and γ, the channel noise coefficient [4]. A review of reported VCOs reveals a wide range in the choice of switching device size and type (NMOS, PMOS, or both). In several cases, the choice of using PMOS transistors is made strictly for their lower $1/f$ noise in that particular technology [8], [9]. We have not seen in the literature any reasons given for choosing an NMOS device or a PMOS device with regard to $1/f^2$ phase noise, largely because the devices have been assumed to yield equivalent thermal noise for equivalent g_m. In Section IV, we show why this assumption does not always hold in short-channel devices. As a result, lower $1/f^2$ phase noise can be achieved using PMOS devices instead of NMOS devices for a given g_m.

Previous work has stated that the tail current source is the primary source of $1/f$ noise and that the contribution due to the MOS cross-coupled pair is made small by the switching action of the oscillator [10]. Ismail *et al.* [11] presented a design that removed the current source and also utilized a suppression technique for the switching transistor $1/f$ noise. In Section V, we will show that reduction in $1/f^3$ phase noise is fundamentally limited by switching transistor $1/f$ noise, and not tail current source $1/f$ noise. We outline a mechanism for $1/f$ noise upconversion and show how to reduce this upconversion factor.

Because phase noise measurements do not provide insight into the relative contributions of circuit noise sources, it is difficult to confirm theories specific to particular noise sources or mechanisms with a single VCO design. In conjunction with simulations, we designed an experimental set of seven VCOs that enabled us to isolate particular noise mechanisms from one another.

Manuscript received December 4, 2003; revised August 15, 2004. This work was supported by the MIT Center for Integrated Circuits and Systems.

The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: ajerng@mit.edu).

Digital Object Identifier 10.1109/JSSC.2004.841035

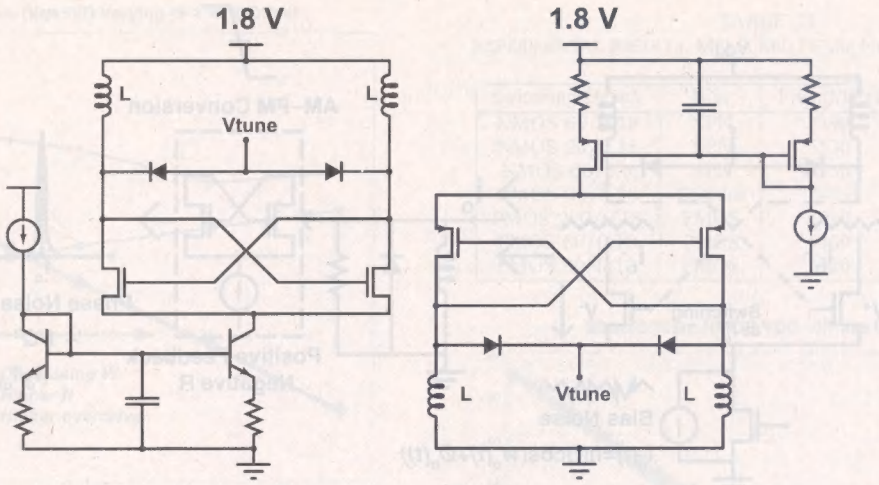


Fig. 1. NMOS and PMOS VCO topologies.

II. VCO Experiment

The VCO topology used in the experiment, shown in Fig. 1, consists of a cross-coupled pair of NMOS or PMOS devices, a tail current source with associated current mirror circuitry, and a differential LC tank that uses standard p^+/n^- junction diodes as varactors. This topology allows low-voltage operation and provides a convenient reference for the varactor to either power or ground, maximizing the voltage tuning range. The VCO circuits were fabricated on a 0.18-μm SiGe BiCMOS process with inductor quality factors (Q) of approximately 10 at 5 GHz.

The phase noise of a VCO is fundamentally related to several key parameters. A semi-empirical model formulated by Leeson expresses this phase noise behavior as [12]

$$\mathcal{L}(f_m) = \frac{2kTR_{\text{eq}}F}{A_o^2}\left(\frac{f_o}{2Qf_m}\right)^2\left(1+\frac{\Delta f_{1/f^3}}{f_m}\right). \quad (1)$$

According to this equation, the parameters that determine phase noise at a frequency offset f_m are the voltage swing, A_o, the tank impedance at resonance, R_{eq}, the tank quality factor, Q, the excess noise factor, F, the $1/f$ corner frequency of the circuit noise, $\Delta f_{1/f^3}$, and the oscillation frequency, f_o. A_o, R_{eq}, Q, and f_o were held constant in our experiment by designing each VCO with the same bias current and tank circuit, allowing meaningful comparison between the noise sources and mechanisms particular to each design. The bias current was set at 7.5 mA to maximize the voltage swing, A_o, under the constraint of gate oxide reliability. In this technology, A_o was limited to 1.8-V_{pk} differential. Key circuit parameters were then varied between the seven VCOs. MOS, NPN, and resistor current sources were implemented, NMOS and PMOS switching transistors were compared, and switching device parameters such as g_m and f_t were varied through device sizing. Finally, the tuning range was kept constant for all designs to maintain a controlled noise term due to AM-PM conversion in the varactor.

A bias current mirror was used to generate the tail current, as opposed to applying an external voltage bias on the tail current transistor. Because we are particularly interested in evaluating the relevance of bias noise, it was essential to not leave out any noise sources which may impact its magnitude. In the basic current mirror configuration with mirror ratio $N > 1$, the output bias current noise is actually dominated by the degeneration resistor in the emitter/source leg of the current mirror device.

III. Bias Noise

A general model shown in Fig. 2 illustrates the conversion of bias noise into phase noise as a two-step process. First, bias current noise $\overline{i_n^2}$ at frequencies ω_n are translated in frequency by the switching action of the MOS cross-coupled pair. Low-frequency bias noise ($\omega_n \ll \omega_o$) mixes up to create two correlated sidebands at $(\omega_o + \omega_n)$ and $(\omega_o - \omega_n)$, resulting in only amplitude modulation (AM) noise. High-frequency bias noise at $\omega_n = (2\omega_o + \Delta w)$ downconverts into a single noise sideband in the passband of the LC tank, containing both AM and PM noise. The resulting output noise current i_o is then amplified and shaped by the positive feedback loop and LC tank filter. This fundamental process of an oscillator limits the output signal and suppresses amplitude noise, implying that only the PM noise components arising from high-frequency bias noise contribute to phase noise. However, the AM noise can potentially be converted into PM noise due to the presence of nonlinear components in the feedback loop.

High-frequency bias noise is attenuated by the low bandwidth of the bias transistors and by the decoupling capacitor at the input to the current mirror. In our design, bias noise near $2\omega_o$, or 10 GHz, is more than an order of magnitude below the level of the low-frequency bias noise. Simulations run with and without a high-frequency bias noise filter yield identical phase noise results. We conclude that high-frequency bias noise is not a significant phase noise contributor to our design.

Amplitude variations due to low-frequency bias noise can be converted into phase noise through modulation of the varactor capacitance. However, this conversion is not fundamental to the design of a VCO. Proper choice of the VCO topology can mitigate varactor AM-PM conversion. Minimizing varactor sensitivity (MHz/V) directly reduces varactor AM-PM conversion at the cost of a reduced tuning range. By adding in parallel to the varactor a bank of digitally switchable capacitors, overall tuning range can be regained without any increase in varactor

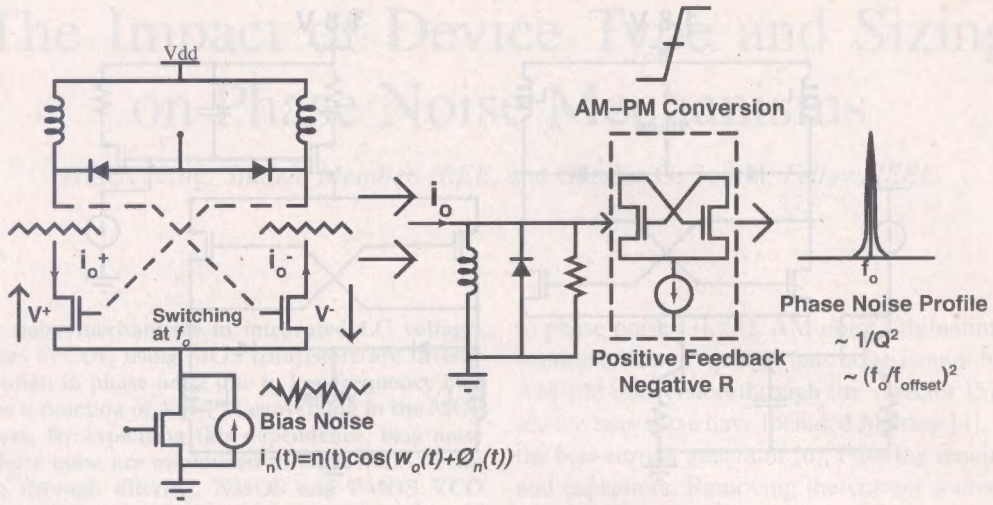


Fig. 2. Bias noise conversion into phase noise.

sensitivity [2]. We designed our VCOs for a tuning range of 400 MHz, corresponding to an average varactor sensitivity of approximately 200 MHz/V. A simulation replacing the varactor with an ideal capacitor showed little change in phase noise, indicating negligible impact of varactor AM-PM conversion on phase noise.

Amplitude variations can also modulate the phase delay associated with the MOS switching devices. The signal swing of an oscillator causes the operating point of the switching transistors to vary periodically. The MOS cross-coupled pair exhibits an AM-PM transfer function that is dependent on its operating characteristics as well as its source and load impedances.

A criteria for oscillation is that the magnitude and phase of the loop gain are unity and 0°, respectively. Phase delays within the loop force an opposite phase shift in the LC tank to maintain 0° phase. As a result, the oscillation frequency ω_o is shifted to [13]

$$\omega_o = \frac{1}{\sqrt{\mathrm{LC}}}\left(1 - \frac{\Delta\theta}{2Q}\right). \tag{2}$$

Table I demonstrates how f_o of a 60 μm/0.18 μm NMOS VCO deviates from its expected ac simulation value of 5690 MHz under large signal conditions. Transient analysis shows that as the amplitude of the oscillation increases through I_{Bias}, f_o shifts downward. The simulations indicate that the frequency shift is related to the level of the harmonic distortion (HD2, HD3) present at the VCO tank output nodes. This effect has been documented and described in [2] as a form of indirect FM. Second and third harmonics of the fundamental current component are generated by the switching transistors. When driven into the LC tank, these harmonics flow into the lower impedance of the capacitance and create an imbalance in reactive power. The oscillation frequency adjusts to compensate for the effective phase shift. Amplitude variations modulate the level of the harmonics, resulting in modulation of the phase shift. Variability in the phase shift results in variability in ω_o, or

TABLE I
DEPENDENCE OF f_o ON I_{Bias}: 60 μm/0.18 μm NMOS VCO

I_{Bias} (mA)	f_o (MHz)	HD2 (dBc)	HD3 (dBc)
1.5	5624	-31	-44
2	5625	-33	-43
4	5518	-24	-36
6	5468	-20	-35
8	5444	-19	-35

phase noise. The phase noise due to this AM-PM mechanism can be expressed as

$$\frac{\mathrm{dBc}}{\mathrm{Hz}}(\Delta\omega) = \frac{1}{2(\Delta\omega)^2}\left|\frac{\partial\omega_o}{\partial I_B}\right|^2 \overline{i_n^2}(\Delta\omega) \tag{3}$$

where $\partial\omega_o/\partial I_B$, the sensitivity of ω_o to bias current fluctuations, and $\overline{i_n^2}(\Delta\omega)$, the magnitude of the bias current noise, can be extracted from ac and transient simulations, respectively. Equation (3) was used to calculate phase noise contributions due to this bias noise mechanism. The calculations matched SpectreRF phase noise simulations.

In order to gain insight into how to minimize the AM-PM conversion factor, we focus on how device sizing can impact VCO harmonics. Increasing the linear range of the MOS differential pair, which is proportional to the gate overdrive $V_{gs} - V_t$, should reduce distortion for a given signal swing. However, at high frequencies, the harmonic distortion is also influenced by device capacitances. Minimizing device capacitance lowers high-frequency distortion. Thus, two convenient metrics for the linearity of the switching transistors are f_t and $V_{gs} - V_t$. These quantities can simultaneously be increased by decreasing the device width of the switching devices since, for a fixed bias current,

$$f_t = \frac{g_m}{C_{gs}} = \frac{\sqrt{2I_D\mu C_{\mathrm{ox}}\frac{W}{L}}}{C_{gs}} \propto \frac{\sqrt{W}}{W} \propto \frac{1}{\sqrt{W}} \tag{4}$$

and

$$V_{gs} - V_t = \sqrt{\frac{I_D L}{\mu C_{\mathrm{ox}} W}} \propto \frac{1}{\sqrt{W}}. \tag{5}$$

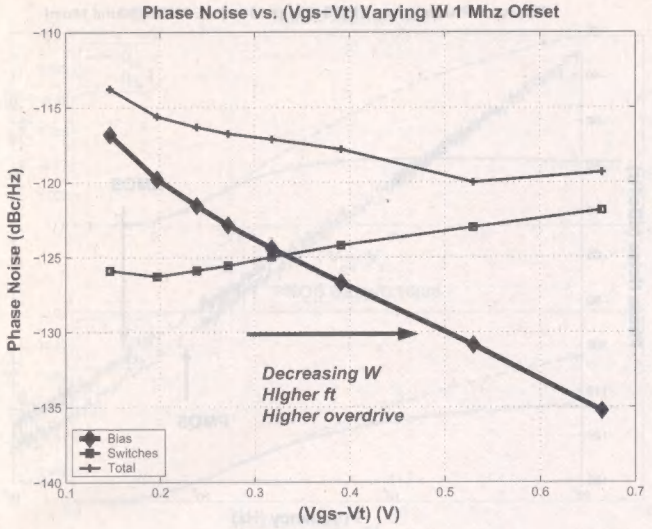


Fig. 3. NMOS VCO phase noise components versus $V_{gs} - V_t$: Vary W.

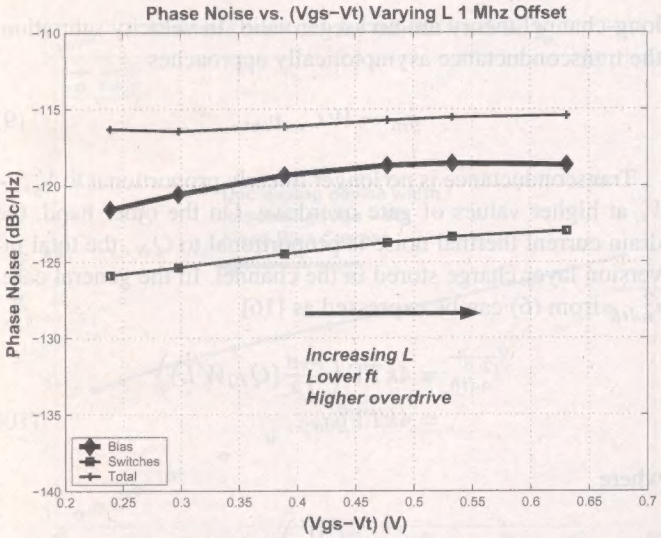


Fig. 4. NMOS VCO phase noise components versus $V_{gs} - V_t$: Vary L.

In short-channel devices, as W becomes small, the dependence of f_t on W becomes weaker and f_t approaches a constant value. The $V_{gs} - V_t$ dependence approaches $1/W$ for small W.

Phase noise simulations using SpectreRF were run to confirm these relationships. We first concentrate on phase noise in the $1/f^2$ region, where flicker noise contributions are small. Figs. 3 and 4 plot simulated NMOS VCO phase noise contributions from bias thermal noise and switching transistor thermal noise at a 1-MHz offset, as well as the total phase noise, as a function of the gate overdrive $V_{gs} - V_t$. Gate overdrive is increased by either reducing device width or increasing device length while keeping a fixed bias current. Reducing device width from 200 to 10 μm reduces the bias noise contribution by 20 dB, making it negligible compared to the noise of the switching transistors and resulting in improvement of the overall phase noise. Increasing channel length from 0.18 to .6 μm increases gate overdrive but reduces f_t, and results in a very slight increase in the bias noise contribution. The results indicate that both the gate overdrive and the f_t of the differential pair are important for linearity.

TABLE II
EXPERIMENTAL RESULTS: MEASURED PHASE NOISE IN $1/f^2$ REGION

Switching Device	Bias	Freq.(MHz)	$\frac{dBc}{Hz}(1\,MHz)$
NMOS 60/0.18	NPN	5390	-114
NMOS 20/0.18	NPN	5530	-118
NMOS 60/0.6	NPN	5350	-113
NMOS 60/0.18	Resistor	5445	-118
PMOS 200/0.18	PMOS	5020	-117
PMOS 60/0.18	PMOS	5309	-122
PMOS 30/0.18	PMOS	5320	-124

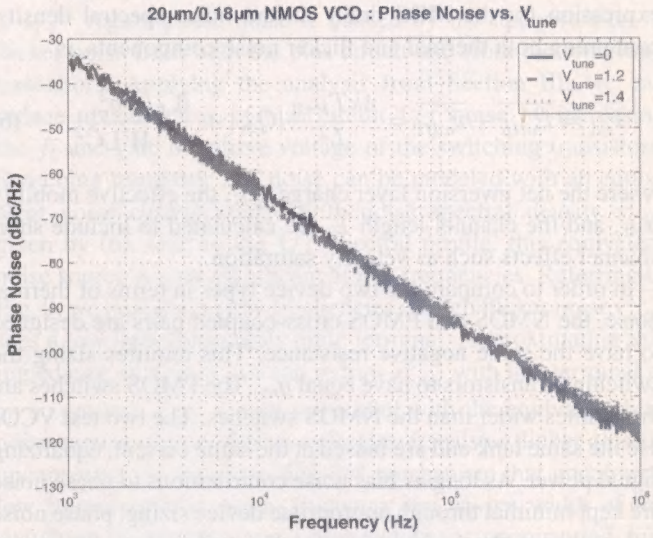


Fig. 5. NMOS VCO phase noise versus V_{tune}.

Notice that bias noise is the primary contributor to $1/f^2$ phase noise in both figures unless minimum length devices with relatively large values of $(V_{gs} - V_t)$ are used.

Table II lists measured phase noise results from the experimental set of VCOs. An NMOS VCO's device width was varied from 60 to 20 μm while a PMOS VCO's device width was varied from 200 to 30 μm. No external capacitors or filtering of any type were applied to any nodes on the bias circuits. Improved phase noise in the $1/f^2$ region (1-MHz offset) is observed for smaller device widths in both NMOS and PMOS designs. In order to prove that this improvement can be attributed to reduced bias noise upconversion, we fabricated an identical 60-μm-width NMOS VCO with its bias circuit replaced with a resistor sized to yield the same bias current. It also showed improved $1/f^2$ phase noise due, in this case, to the replacement of the current mirror bias noise with the much smaller noise contribution of a single resistor. An NMOS VCO with its length scaled from 0.18 to 0.6 μm had slightly degraded $1/f^2$ phase noise as expected. Finally, Fig. 5 plots measured phase noise for the 20-μm-width NMOS VCO at varactor tuning voltages ranging from 0 to 1.4 V. The three curves show little difference, even though the varactor tuning sensitivity is varying by a factor of 3 over this range. The impact of varactor AM-PM conversion on phase noise is minor.

The degree of width reduction required for adequate bias noise suppression is a function of how noisy the bias circuit is. From a practical standpoint, reduction of device width is limited by two constraints. First, the loop gain of the VCO must be sufficiently greater than 1 to guarantee oscillation over

all operating conditions, leading to a minimum required value for g_m. Second, the increase in $V_{gs} - V_t$ is limited by the voltage headroom required by the current source, and also by the supply voltage being used.

IV. MOS Channel Thermal Noise

After reducing bias noise contributions, MOS channel thermal noise in the NMOS or PMOS switching transistors is the main source of phase noise in the $1/f^2$ region. An expression for the MOS drain current noise spectral density, containing both thermal and flicker noise components, is

$$\overline{i_{\rm nd}^2} = \overline{i_{ndth}^2} + \overline{i_{ndfl}^2} = \frac{4kT\mu_{\rm eff}}{L^2}|Q_N| + \frac{K_f}{f}\frac{g_m^2}{WLC_{\rm ox}^2} \quad (6)$$

where the net inversion layer charge Q_N, the effective mobility $\mu_{\rm eff}$, and the channel length L, are calculated to include short channel effects such as velocity saturation.

In order to compare the two device types in terms of thermal noise, the NMOS and PMOS cross-coupled pairs are designed to have the same negative resistance. This requires sizing the switching transistors to have equal g_m. The PMOS switches are three times wider than the NMOS switches. The two test VCOs use the same tank and are biased at the same current, equalizing signal power. As long as bias noise contributions to phase noise are kept minimal through appropriate device sizing, phase noise differences in the $1/f^2$ region can be attributed to differences in the drain current thermal noise between NMOS or PMOS switching transistors.

Using first-order long-channel MOS theory, we can express the drain current thermal noise as

$$\overline{i_{ndth}^2} = 4kT\left(\frac{2}{3}g_m\right) \quad (7)$$

According to (7), the NMOS and PMOS VCOs should exhibit the same phase noise performance in the $1/f^2$ region.

Fig. 6 plots measured phase noise for NMOS and PMOS VCOs with device dimensions of 20 μm/.18 μm and 60 μm/.18 μm, respectively. The PMOS VCO has ~4 dB better phase noise than the NMOS VCO in the $1/f^2$ region, indicating lower drain current thermal noise in the PMOS device. In order to understand this, the effects of velocity saturation in short-channel devices must be considered. The carrier velocity is a function of the horizontal electric field in the channel and can be modeled by the following piecewise equation [14], [15]:

$$\begin{aligned} v &= \frac{\mu_{\rm eff}E}{1+E/E_C}, \qquad E \le E_C \\ &= v_{\rm sat}, \qquad E > E_C \end{aligned} \quad (8)$$

where E_C is the critical field at which the carriers are velocity saturated and is equal to $2v_{\rm sat}/\mu_{\rm eff}$.

NMOS devices suffer from velocity saturation more than PMOS devices because their critical electric field, E_C, is much lower [16]. Velocity saturation is important when the channel length, L, is small, and the gate overdrive voltage, $V_{gs} - V_t$, is high. When $V_{gs} - V_t$ approaches the product LE_c, the equations for drain current and transconductance based on long-channel theory are no longer valid. In velocity saturation, the transconductance asymptotically approaches

$$g_m = WC_{\rm ox}v_{\rm sat}. \quad (9)$$

Transconductance is no longer linearly proportional to $V_{gs} - V_t$ at higher values of gate overdrive. On the other hand, the drain current thermal noise is proportional to Q_N, the total inversion layer charge stored in the channel. In the general case, $\overline{i_{ndth}^2}$ from (6) can be expressed as [16]

$$\begin{aligned} \overline{i_{ndth}^2} &= 4kT\Gamma\left(\frac{\mu_{\rm eff}}{L^2}(Q_{I0}WL)\right) \\ &= 4kT\Gamma g_{do} \end{aligned} \quad (10)$$

where

$$g_{do} = \mu_{\rm eff}C_{\rm ox}\left(\frac{W}{L}\right)(V_{gs} - V_t). \quad (11)$$

Γ is a function of the product LE_c, and bias conditions V_{ds}, and $V_{gs} - V_t$. According to (10) and (11), $\overline{i_{ndth}^2}$ remains proportional to $V_{gs} - V_t$. The amount of drain current thermal noise for a given g_m can thus be related to the ratio g_{do}/g_m. This ratio is equal to one for low $V_{gs} - V_t$ but increases as the device enters velocity saturation. Fig. 7 plots measured g_m and g_{do} curves versus V_{gs} for the NMOS and PMOS device sizes used in the test VCOs. The g_{do}/g_m ratio is clearly greater in the NMOS device for most gate bias voltages.

Simulations using BSIM3 v3 models confirm the degrading effects of velocity saturation. Fig. 8 plots $\overline{i_{ndth}^2}$ of 0.18-μm NMOS and PMOS transistors versus g_m for a fixed bias current. g_m is varied by changing the device width. On this plot, lower values of g_m correspond to using smaller width devices biased at higher gate overdrive. The NMOS devices show higher output current noise for the same g_m, with the difference becoming greater as g_m becomes smaller, or as $V_{gs} - V_t$ becomes larger. The devices in our two test VCOs operate at a g_m of approximately 10 mS, where the ratio between NMOS and PMOS drain current thermal noise is about 2.3.

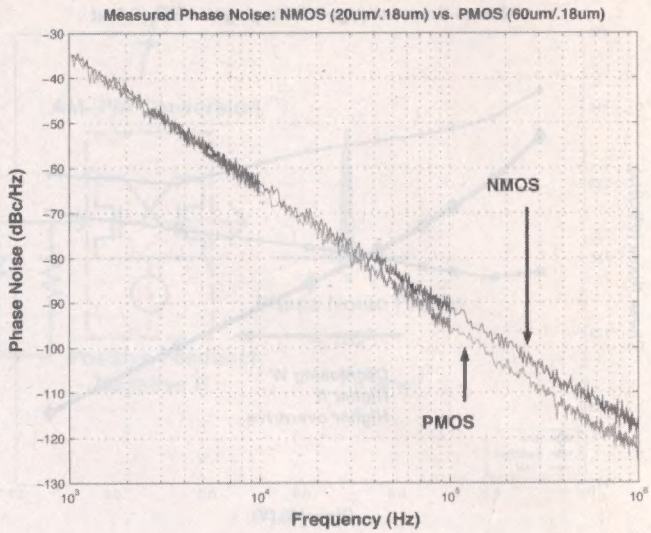


Fig. 6. Phase noise: NMOS (20μm/0.18μm) versus PMOS (60μm/0.18μm).

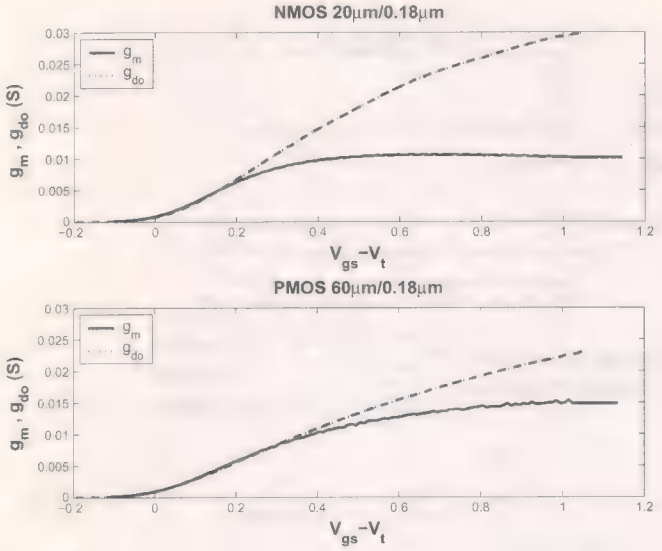


Fig. 7. Measured g_m and g_{do} versus V_{gs} for 0.18-μm NMOS and PMOS.

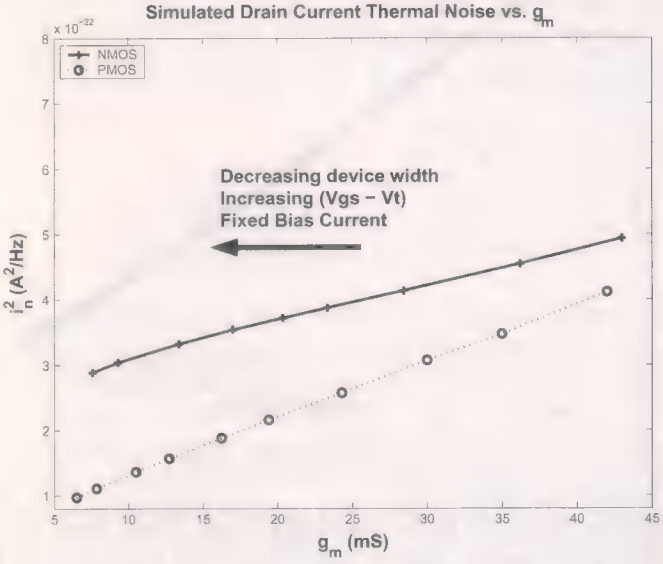


Fig. 8. Simulated $\overline{i_{ndth}^2}$ versus g_m for 0.18-μm NMOS and PMOS.

Phase noise simulations indicate that the drain current thermal noise contribution of the switching transistors is about 4 dB less in the PMOS VCO than in the NMOS VCO, agreeing well with measured results. In order to evaluate this difference more fairly, we must consider an additional phase noise dependency. In this particular VCO topology, the Q of the NMOS tank is slightly lower than the Q of the PMOS tank. As shown in Fig. 1, the parasitic diode on the varactor cathode directly loads the NMOS tank. In the PMOS VCO, the parasitic diode is on a virtual ground and has no effect on Q. Since phase noise is proportional to $1/Q^2$, a degradation of 2 dB in the NMOS VCO is expected according to tank Q simulations. This suggests that the actual difference due to drain current thermal noise alone is about 2 dB. In order to confirm this, a simulation where the varactor was replaced with an ideal capacitor was run for both VCOs. The results showed a 2.3-dB difference between NMOS and PMOS drain current thermal noise contributions to phase noise.

The improved phase noise performance derived from using PMOS switching transistors is a byproduct of the need to suppress bias noise contributions. High current densities are required in the switches to reduce the bias noise below the level of the switching device thermal noise. Under these bias conditions, VCOs using PMOS switches achieve significantly lower phase noise when using deep-submicron CMOS.

V. MOS Flicker Noise

$1/f^3$ region phase noise is caused by the upconversion of flicker noise from both the bias circuit and from the switching transistors. Applying the analysis from Section III, we can reduce upconversion of bias circuit $1/f$ noise by increasing the f_t and gate overdrive voltage of the switching transistors. Switching transistor $1/f$ noise can be modeled with an equivalent noise current source with noise spectral density $\overline{i_{ndfl}^2}$ given by (6). Due to the $1/f$ spectral profile, this equivalent noise source is only significant at low frequencies. Referring to Fig. 9, we note that at low frequencies, switch flicker noise $\overline{i_{ndfl}^2}$ sees a low impedance at its drain terminal. Approximating this impedance as a short, we can redraw $\overline{i_{ndfl}^2}$ with this terminal at an ac ground. $\overline{i_{ndfl}^2}$ is now in parallel with the equivalent bias noise current source. Hence, switching transistor flicker noise is upconverted via the same AM-PM mechanism that upconverts low frequency bias noise. Reducing the device width of the switching transistors should also reduce its upconverted $1/f$ noise. However, there are two differences between the $1/f$ bias noise and the $1/f$ switching transistor noise. First, changing the size of the switching transistors affects the magnitude of $\overline{i_{ndfl}^2}$ but does not affect the magnitude of the bias noise. Second, unlike the bias noise which is stationary, $\overline{i_{ndfl}^2}$ depends on the operating point of the switching transistors and varies periodically as a function of time.

Fig. 10 plots simulated phase noise contributions at a 10-kHz offset as a function of $V_{gs} - V_t$ while varying the device width of a PMOS VCO. While the $1/f$ bias noise contribution drops rapidly as $V_{gs} - V_t$ increases, the $1/f$ switching transistor noise contribution decreases more gradually, indicating additional dependencies. Optimization of the total phase noise at a 10-kHz offset requires the use of small device widths, in which case the flicker noise contribution from the switching transistors dominates over that from the bias. In the case of the bias flicker noise, one can simultaneously reduce both the magnitude and the upconversion factor of its noise, since the former involves sizing of the bias transistors while the latter involves sizing of the switching transistors. The current source transistor used in our PMOS design had device dimensions of 2000 μm/1 μm. On the other hand, sizing of the switching transistors involves a tradeoff between the $1/f$ noise magnitude and the $1/f$ noise upconversion factor.

Experimental results confirm that in both NMOS and PMOS VCOs, reducing device width improves phase noise in the $1/f^3$ region (Table III). The best phase noise of -70 dBc/Hz at a 10-kHz offset is achieved by the 30 μm/0.18 μm PMOS VCO. In this technology, the PMOS devices have lower $1/f$ noise than the NMOS devices. Our experiment also compares the relative contributions to $1/f^3$ phase noise between the bias transistors

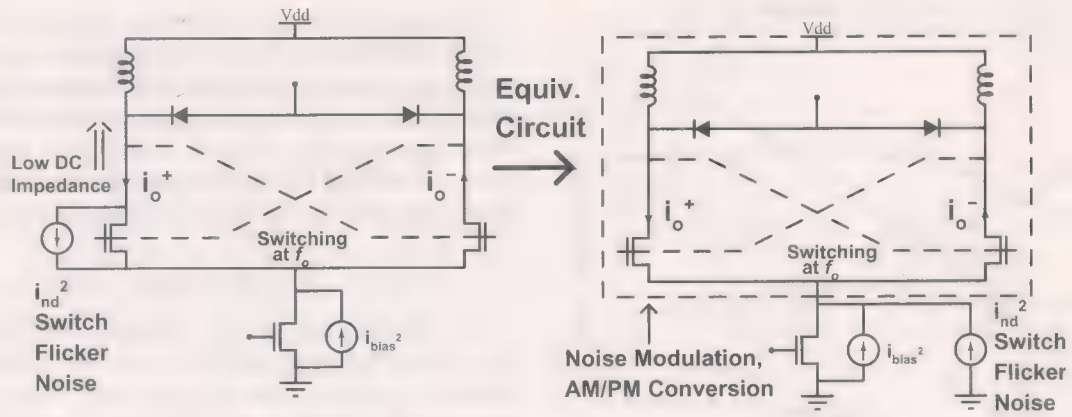


Fig. 9. Switching transistor $1/f$ noise: Upconversion model.

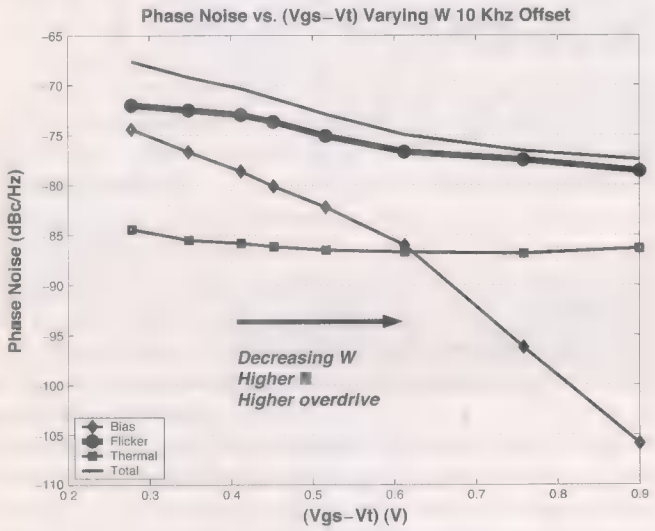


Fig. 10. PMOS VCO phase noise components versus $(V_{gs} - V_t)$: Vary W.

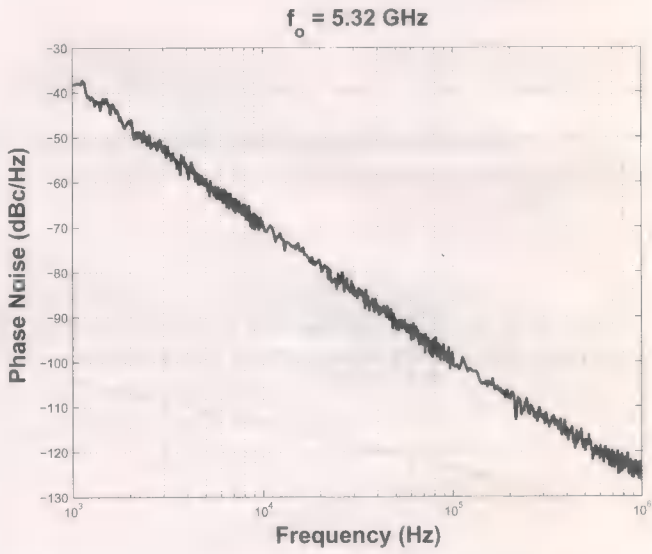


Fig. 11. Measured phase noise: 30 μm/0.18 μm PMOS VCO.

TABLE III
EXPERIMENTAL RESULTS: MEASURED PHASE NOISE IN $1/f^3$ REGION

Switching Device	Bias	Freq.(MHz)	$\frac{dBc}{Hz}(10\,kHz)$
NMOS 60/0.18	NPN	5390	-59
NMOS 20/0.18	NPN	5530	-65
NMOS 60/0.6	NPN	5350	-60
NMOS 60/0.18	NMOS	5390	-58
NMOS 60/0.18	Resistor	5445	-59
PMOS 200/0.18	PMOS	5020	-63
PMOS 60/0.18	PMOS	5309	-65
PMOS 30/0.18	PMOS	5320	-70

and the switching transistors. The 60 μm/0.18 μm NMOS VCO was designed with an NMOS current source, and with NPN and resistor current sources that have virtually no flicker noise. At a 10-kHz offset, the phase noise of the three VCOs is similar, indicating that the dominant contributor to $1/f^3$ phase noise is the $1/f$ noise of the switching transistors.

The importance of the $1/f$ noise upconversion factor is evident when comparing the 60 μm/0.6 μm NMOS and 60 μm/0.18 μm PMOS VCOs to the 20 μm/0.18 μm NMOS design. Although the 0.6-μm-length NMOS device should have less flicker noise than the 20 μm/0.18 μm device, it has 5 dB worse phase noise at a 10-kHz offset. Likewise, the models indicate that the PMOS devices have less flicker noise than the NMOS devices. Instead, the 60 μm/0.18 μm PMOS and 20 μm/0.18 μm NMOS VCOs have the same phase noise at a 10-kHz offset. The higher f_t of the 20 μm/0.18 μm NMOS device reduces AM-PM conversion and lowers the $1/f$ upconversion factor.

VI. OPTIMIZED ALL-PMOS VCO TOPOLOGY

Analysis of phase noise mechanisms has shown why PMOS switching devices provides better phase noise performance than NMOS devices in both the $1/f^3$ **and** $1/f^2$ regions in a 0.18-μm $L_{\min}$ technology. Fig. 11 shows the measured phase noise for the optimally sized 30 μm/0.18 μm PMOS design. Phase noise of -124 dBc/Hz is achieved at a 1-MHz offset and a center frequency of 5.32 GHz. The design operates from a 1.8-V supply and consumes 7.5 mA of bias current. For a tuning voltage range from 0 to 1.8 V, the VCO tunes 400 MHz, or approximately 8%.

The all-PMOS VCO circuit also offers several advantages from a topology standpoint. It provides excellent isolation from power supply noise through the use of a PMOS current source to VDD, and a ground-referenced tank. Fig. 12 illustrates the effect of power supply noise on the all-PMOS topology in contrast

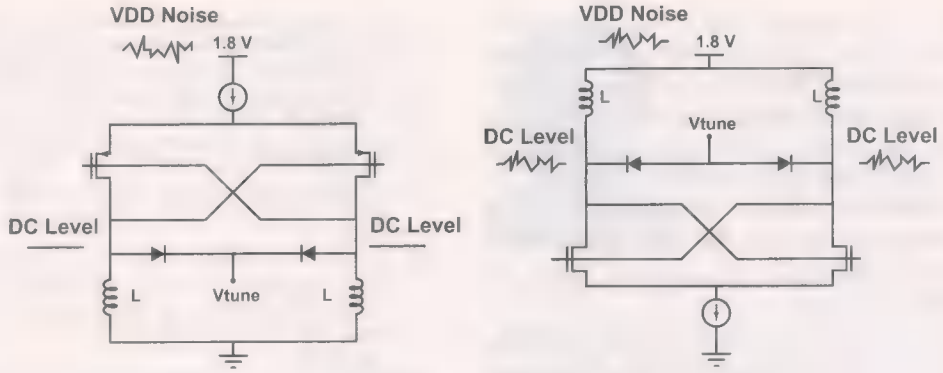


Fig. 12. Effect of VDD noise on VCO topologies.

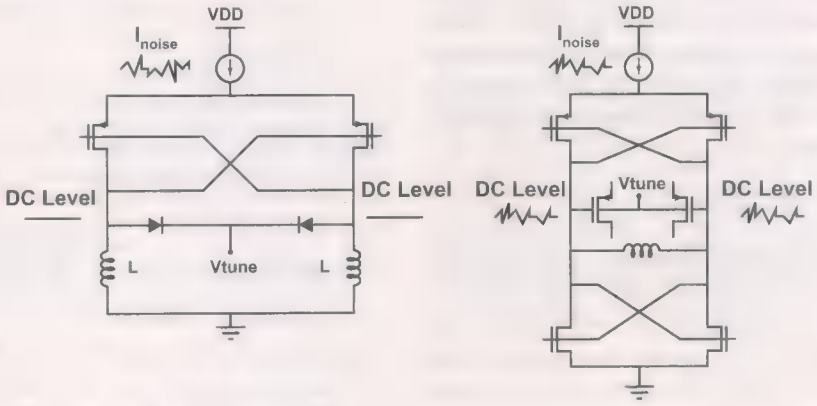


Fig. 13. Effect of I_{bias} noise on varactor dc bias.

to an NMOS topology. The NMOS circuit allows supply noise to couple into the oscillator feedback loop. In addition, low frequency supply noise directly modulates the voltage across the varactor in the NMOS case, inducing phase noise through FM modulation. A supply noise rejection simulation is run using the PXF analysis in SpectreRF. The periodic transfer function for low frequency noise from VDD to the VCO output nodes is 20 dB less in the PMOS topology than in the NMOS topology. Finally, if one uses p^+/n^- junction varactor diodes, the PMOS tank Q will be higher, as discussed in Section IV.

The all-PMOS topology can scale down to lower supply voltages than the double cross-coupled topology, which uses NMOS and PMOS differential pairs and requires an additional V_{gs} of voltage headroom. The extra V_{gs} makes it difficult to bias the switching devices at a high gate overdrive for optimized phase noise. Fig. 13 compares the effect of bias current noise on the dc level of the varactors in the two topologies. Noise fluctuations on the bias current modulate the V_{gs} of the bottom NMOS transistors in the double cross-coupled topology. The dc bias on the varactors varies, resulting in modulation of the varactor capacitance. In the all-PMOS topology, bias current noise can modulate the V_{gs} of the PMOS switching pair but does not change the dc bias point of the varactors, which are referenced to ground through a low dc impedance.

The ground-referenced tank serves to minimize noise disturbances to the varactor, allowing the achievement of higher values of K_{vco} without degradation in phase noise. In summary, the all-PMOS VCO topology is desirable because it minimizes both intrinsic and extrinsic sources of phase noise.

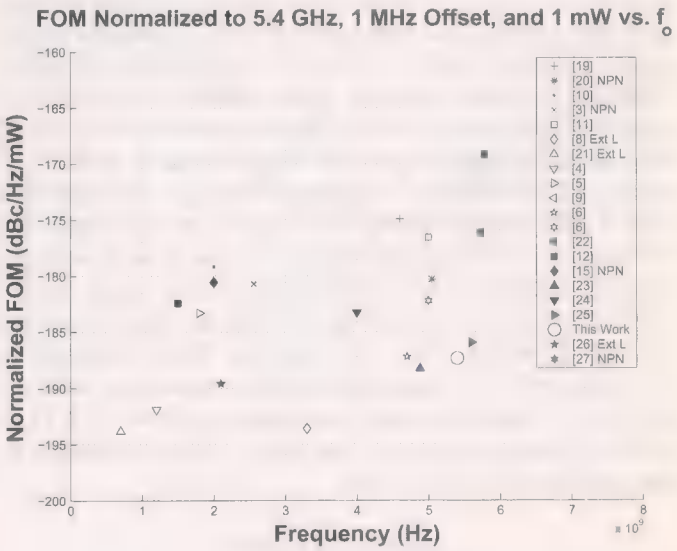


Fig. 14. Comparison of normalized figure of merit versus f_o.

VII. Comparison of Results

In comparing our phase noise results to other published results, we normalize all phase noise data to a center frequency of 5.4 GHz and a frequency offset of 1 MHz. Fig. 14 graphs the normalized phase noise figure of merit (FOM) of this work and other published data against center frequency f_o [3]–[6], [8]–[11], [17]–[27]. The graph shows a general upward trend, with the normalized FOM degrading as the oscillation frequency

increases. Our work lies at the bottom right corner of the graph, demonstrating excellent phase noise at a high oscillation frequency. Of the results with better phase noise performance, three use high Q external inductors while another one operates at a much lower center frequency of 1.2 GHz. In light of this paper's analysis on AM-PM conversion in the switching transistors, we postulate that the degradation of phase noise performance at higher oscillation frequencies seen in this graph is due to an increase in low-frequency bias noise upconversion, which our design has specifically minimized.

We emphasize that our results are achieved without adding additional on-chip LC bias filters or external decoupling capacitors, maintaining the use of a standard current mirror and current source. Finally, although biasing at 7.5 mA optimizes our phase noise, it does not necessarily optimize our FOM. We can lower our power consumption by scaling down the bias current to 1 mA. At this bias condition, the measured phase noise at a 1-MHz offset is -118 dBc/Hz, corresponding to an optimized FOM of -190 dBc/Hz/mW.

VIII. Conclusion

Several new concepts are proposed for the optimization of phase noise. Switching transistor device width should be minimized to lower the upconversion factor for low frequency bias noise and switching transistor flicker noise. An important benefit of this is that bias noise contributions to phase noise are minimized without needing to add filters or remove the current source. The fact that $1/f^3$ phase noise is improved through a reduction in the size of the switching transistors is counter-intuitive and highlights the importance of the upconversion factor. PMOS switching transistors should be used instead of NMOS switching transistors because, under optimal bias conditions, they contribute less drain current thermal noise for the same g_m. This results in improved phase noise performance in the $1/f^2$ region. An all-PMOS VCO topology using a ground referenced tank is effective in reducing the influence of secondary noise mechanisms such as the upconversion of bias noise through the varactor, and the upconversion of low-frequency supply noise.

Key dependencies between phase noise and device parameters are derived from theoretical analysis. These relationships are confirmed in both simulations and measured phase noise results taken from a systematic experiment consisting of 8 VCO designs. Proper device choice and device sizing are essential to the optimization of phase noise.

Acknowledgment

The authors would like to thank J. Gross at IBM Microelectronics for chip fabrication, and T. Sepke for his invaluable discussions on device noise issues and for providing measured data on the 0.18-μm devices.

References

[1] A. Hajimiri and T. H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[2] J. Rael and A. Abidi, "Physical processes of phase noise in differential LC oscillators," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, 2000, pp. 569–572.

[3] A. Zanchi, C. Samori, S. Levantino, and A. L. Lacaita, "A 2-V 2.5-GHz–104-dBc/Hz at 100 kHz fully integrated VCO with wide-band low-noise automatic amplitude control loop," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 611–619, Apr. 2001.

[4] E. Hegazi, H. Sjoland, and A. A. Abidi, "A filtering technique to lower LC oscillator phase noise," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 1921–1930, Dec. 2001.

[5] P. Andreani and H. Sjoland, "Tail current noise suppression in RF CMOS VCOs," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 342–348, Mar. 2002.

[6] S. Levantino, C. Samori, A. Bonfanti, S. L. J. Gierkink, A. L. Lacaita, and V. Boccuzzi, "Frequency dependence on bias current in 5-GHz CMOS VCOs: Impact on tuning range and flicker noise upconversion," *IEEE J. Solid-State Circuits*, vol. 37, no. 8, pp. 1003–1011, Aug. 2002.

[7] E. Hegazi and A. A. Abidi, "Varactor characteristics, oscillator tuning curves, and AM-FM conversion," *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 1033–1039, Jun. 2003.

[8] J. Kucera and B.-U. Klepser, "3.6 GHz VCO's for multi-band GSM transceivers," in *Proc. Eur. Solid-State Circuits Conf. (ESSCIRC)*, 2002, pp. 201–204.

[9] C.-M. Hung, B. A. Floyd, and K. K. O. XQXQXQ, "A fully integrated 5.35-GHz CMOS VCO and a prescalar," in *Proc. IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, 2000, pp. 69–72.

[10] B. D. Muer, M. Borremans, M. Steyaert, and G. L. Puma, "A 2-GHz low-phase-noise integrated LC-VCO set with flicker-noise upconversion minimization," *IEEE J. Solid-State Circuits*, vol. 35, no. 7, pp. 1034–1038, Jul. 2000.

[11] A. Ismail and A. A. Abidi, "CMOS differential LC oscillator with suppressed up-converted flicker noise," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2003, pp. 98–99.

[12] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[13] C. Samori, A. L. Lacaita, A. Zanchi, S. Levantinoi, and G. Cali, "Phase noise degradation at high oscillation amplitudes in LC-tuned VCO's," *IEEE J. Solid-State Circuits*, vol. 35, no. 1, pp. 96–99, Jan. 2000.

[14] B. Hoefflinger, H. Sibbert, and G. Zimmer, "Model and performance of hot-electron MOS transistor for VLSI," *IEEE Trans. Electron Devices*, vol. ED-26, no. 4, p. 513, Apr. 1979.

[15] C. G. Sodini, P.-K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. ED-31, no. 10, pp. 1386–1393, Oct. 1984.

[16] B. Wang, J. R. Hellums, and C. G. Sodini, "MOSFET thermal noise modeling for analog integrated circuits," *IEEE J. Solid-State Circuits*, vol. 29, no. 7, pp. 833–835, Jul. 1994.

[17] P. Kinget, "A fully integrated 2.7 V 0.35 um CMOS LC VCO for 5 GHz wireless applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 1998, pp. 226–227.

[18] M. S. J.-O. Plouchart, H. Ainspan, and A. Ruehli, "A fully-monolithic SiGe differential voltage-controlled oscillator for 5 GHz wireless applications," in *Proc. IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, 2000, pp. 57–60.

[19] C. Samori, S. Levantino, and V. Boccuzzi, "A -94 dBc/Hz at 100 kHz, fully-integrated, 5-GHz, CMOS VCO with 18% tuning range for Bluetooth applications," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, 2001, pp. 201–204.

[20] K. Hoshino, E. Hegazi, J. Rael, and A. Abidi, "A 1.5 V 1.7 mA 700 MHz CMOS LC oscillator with no upconverted flicker noise," in *Proc. Eur. Solid-State Circuits Conf. (ESSCIRC)*, Sep. 2001.

[21] Y. Lin, K. To, J. Hamel, and W. Huang, "Fully integrated 5 GHz CMOS VCO's with on chip low frequency feedback circuit for $1/f$ induced phase noise suppression," in *Proc. Eur. Solid-State Circuits Conf. (ESSCIRC)*, 2002.

[22] T. K. Johansen and L. E. Larson, "Optimization of SiGe VCO's for wireless applications," in *Proc. IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, 2002, pp. 201–204.

[23] S. L. Gierkink, S. Levantino, R. C. Frye, C. Samori, and V. Boccuzzi, "A low-phase-noise 5-GHz CMOS quadrature VCO using super-harmonic coupling," *IEEE J. Solid-State Circuits*, vol. 38, no. 7, pp. 1148–1154, Jul. 2003.

[24] J. Maget, M. Tiebout, and R. Kraus, "MOS varactors with n and p-type gates and their influence on an LC-VCO in digital CMOS," *IEEE J. Solid-State Circuits*, vol. 38, no. 7, pp. 1139–1147, Jul. 2003.

[25] N. H. Fong, J.-O. Plouchart, N. Zambmer, D. Liu, L. F. Wagner, C. Plett, and N. G. Tarr, "Design of wide-band CMOS VCO for multiband wireless LAN applications," *IEEE J. Solid-State Circuits*, vol. 38, no. 8, pp. 1333–1342, Aug. 2003.

[26] J. J. Kucera, "Wideband BiCMOS VCO for GSM/UMTS direct conversion receivers," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2001, pp. 374–375.

[27] G. Grau, U. Langman, W. Winkler, D. Knoll, J. Osten, and K. Pressel, "A current-folded up-conversion mixer and VCO with center-tapped inductor in a SiGe-HBT technology for 5-GHz wireless LAN applications," *IEEE J. Solid-State Circuits*, vol. 35, no. 9, pp. 1345–1351, Sep. 2000.

Albert Jerng (M'97–S'02) was born in Princeton, NJ, in 1972. He received the B.S.E.E. and M.S.E.E. degrees from Stanford University, Stanford, CA, in 1994 and 1996, respectively. He is currently pursuing the Ph.D. degree at the Massachusetts Institute of Technology, Cambridge, where his research interests include low phase noise VCOs and transmit DACs for high data rate wireless transceivers.

He was with Advanced Micro Devices from 1996 to 1999 designing RF integrated circuits for cellular and cordless phone applications. From 1999 to 2002, he worked at DSP Group, where he developed CMOS RF transceivers for 900-MHz and 2.4-GHz cordless phone systems.

Mr. Jerng served on the IEEE RFIC Symposium Steering Committee during 2003–2004.

Charles G. Sodini (S'80–M'82–SM'90–F'94) was born in Pittsburgh, PA, in 1952. He received the B.S.E.E. degree from Purdue University, Lafayette, IN, in 1974, and the M.S.E.E. and the Ph.D. degrees from the University of California, Berkeley, in 1981 and 1982, respectively.

He was a Member of the Technical Staff at Hewlett-Packard Laboratories from 1974 to 1982, where he worked on the design of MOS memory and later, on the development of MOS devices with very thin gate dielectrics. He joined the faculty of the Massachusetts Institute of Technology, Cambridge, in 1983, where he is currently a Professor in the Department of Electrical Engineering and Computer Science. His research interests are focused on integrated circuit and system design with emphasis on analog, RF and memory circuits and systems. Along with Prof. Roger T. Howe, he is a co-author of an undergraduate text on integrated circuits and devices entitled *Microelectronics: An Integrated Approach*. He also studied the Hong Kong electronics industry and co-authored a chapter with Prof. Rafael Reif in a recent book entitled *Made by Hong Kong*.

Dr. Sodini held the Analog Devices Career Development Professorship of Massachusetts Institute of Technology's Department of Electrical Engineering and Computer Science and was awarded the IBM Faculty Development Award from 1985 to 1987. He was the Associate Director of MIT's Microsystems Technology Laboratories from 1989 to 1996. He has served on a variety of IEEE Conference Committees, including the International Electron Device Meeting, where he was the 1989 General Chairman. He was the Technical Program Co-Chairman for the 1992 Symposium on VLSI Circuits and the 1993–1994 Co-Chairman of the Symposium. He served on the Electron Device Society Administrative Committee from 1988–1994. He is the past president of the IEEE Solid-State Circuits Society and a member of its Administrative Committee.

Analysis and Simulation of Spectral Regrowth in Radio Frequency Power Amplifiers

Burcin Baytekin, *Student Member, IEEE,* and Robert G. Meyer, *Fellow, IEEE*

***Abstract*—This paper presents a novel method for efficiently analyzing the relationship between spectral regrowth and physical distortion mechanisms in radio frequency power amplifiers. It utilizes a Volterra series model whose coefficients are computed from basic SPICE parameters. The analysis uses a decomposition of the Volterra kernels into simpler subsystems in order to greatly reduce the computation times. The method is applied to the design of several bipolar-transistor power amplifiers after a series-based model is developed for representing the increase in active device forward transit time at high collector current densities. A number of single-stage SiGe power amplifiers have been designed, fabricated, and tested using the IEEE802.11b and IS-95 modulation schemes at different carrier frequencies, and these results are compared with the theoretical analysis.**

***Index Terms*—Adjacent channel power ratio (ACPR), distortion, forward transit time, power amplifiers, spectral regrowth, Volterra series.**

I. Introduction

LINEAR power amplifiers (PAs) are becoming widely used in wireless communication systems with the rising popularity of code-division multiple access (CDMA) and orthogonal frequency-division multiplex (OFDM) systems. The envelope of the signals in these communications systems are not constant, so that the PA design for these systems must pay attention to the sources of nonlinearity in the PA in order to limit the amount of spectral regrowth, which can cause unacceptable levels of interference in the adjacent channels.

A power amplifier has to supply all of the radiated power at the transmitter antenna, as well as the power lost through the passive elements such as radio frequency (RF) filters or duplexers. This makes the PA efficiency the dominant factor in the total power dissipation of the radio transmitter, which is especially significant for mobile communication applications.

The trade-off between efficiency and linearity leads PA designers to search for an optimum device operating point. Therefore, a good understanding of the effects of the transistor components and PA design parameters on linearity is essential. Linearity has long been analyzed by using intermodulation distortion (IM_3) as a metric, instead of the transmit spectrum mask, adjacent channel power ratio (ACPR), or error vector magnitude (EVM) specifications used in the wireless standards. The phenomenon of spectral regrowth has been analyzed in recent publications, but these treatments employ empirical methods which require curve-fitting a function (such as a real or complex baseband-equivalent power series) to the AM-AM and AM-PM simulations or measurements [1]–[7]. This approach does not provide much insight into the relationship between the physical mechanisms in the circuit and spectral regrowth. Some of these analyses also assume that the input signals have a Gaussian amplitude distribution, although all of the conventional digital communications systems are based on the transmission of a set of discrete symbols with equal probability.

As the carrier frequency in wireless systems can be 2 to 4 orders of magnitude greater than the bandwidth of the signal, simulating the performance of PAs with properly modulated signals is impractical with the conventional time-domain methods. This leads designers to using rule-of-thumb methods involving single or two-tone test simulations, although there is no simple relationship between the results of these tests and ACPR type specifications.

In this paper, a novel method is proposed to predict spectral regrowth in PAs. The method uses basic SPICE parameters, which are based on the active device physical mechanisms, in order to make it applicable to transistors fabricated by different processes. First, a Volterra series model of the PA is calculated from the SPICE parameters, and the spectral regrowth is then predicted by using modulated signals. The decomposition of the Volterra kernels into simpler subsystems as proposed in Section III allows the combination of frequency and time-domain computations so that numerical results can be rapidly calculated. These results can assist circuit designers in understanding the effect of design parameters on spectral regrowth and the trade-off between efficiency and linearity. A better understanding of these issues allows them to determine their initial design parameters more accurately before they initiate detailed simulations, helping them avoid time-consuming iterations. Identifying the transistor components contributing to distortion helps device designers optimize the power transistors.

Section II of this paper presents a brief overview of the Volterra series and Section III explains how to obtain numerical results from the analysis. Section IV presents an implementation example and Section V provides the results obtained, while Section VI concludes the paper.

Manuscript received April 6, 2004; revised August 23, 2004. This work was supported by the U.S. Army Research Office under Grant DAAD19-00-1-0550.

B. Baytekin was with the Electronics Research Laboratory, Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA. He is now with Sequoia Communications, San Diego, CA 92127 USA (e-mail: baytekin@cal.berkeley.edu)

R. G. Meyer is with the Electronics Research Laboratory, Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA.

Digital Object Identifier 10.1109/JSSC.2004.840968

II. Volterra Series

If a nonlinear system does not have memory, the output can be expressed as a Taylor series

$$y(x) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{1}$$

which models weakly nonlinear behavior reasonably well. However, RF power amplifiers include circuit elements, such as capacitors and inductors, whose impedances vary with frequency. This variation introduces memory into the system, which may be modeled by means of a Volterra series, as shown in (2) at the bottom of the page, where the functions $h(\tau_1, \tau_2, \ldots, \tau_n)$ are the Volterra kernels of the system [8].

If a causal system described by a Volterra series lacks memory, then

$$h(\tau_1, \tau_2, \ldots, \tau_n) = 0, \text{ for any } \tau_j \neq 0, \ j = 1, 2, \ldots, n \tag{3}$$

and (2) reduces to a Taylor series.

In narrowband systems, the distortion products due to even-order kernels fall into frequency bands well removed from the desired signal as shown in the Appendix [9]. Hence, even-order Volterra kernels can be neglected for the analysis of spectral regrowth. Volterra kernels in (2) beyond the third are usually neglected in order to make a practical evaluation of $y(t)$ possible. Although inclusion of the higher order terms would result in a more accurate representation, the accuracy of these kernels depend on the accuracy of the derivatives of the nonlinear functions in the circuit, which are difficult to determine precisely. In practice, useful results are obtained neglecting terms beyond third order and this approach is followed here.

III. Method of Computation

A. Time and Frequency-Domain Calculations

The input $x(t)$ used in a typical communication system does not take on deterministic values, but is composed of a signal modulated by random data and a specified modulation scheme, requiring the generation of large number of bits for a proper simulation. Furthermore, the carrier frequency in wireless communication systems can be 2 to 4 orders of magnitude larger than the bandwidth of the information-carrying signal or the envelope. Therefore, computation in the time domain requires too many samples per bit of data for practical circuit simulations. Frequency-domain calculations require far fewer samples, because the computation can be limited to the frequency bands of interest in narrowband systems. Furthermore, taking the Fourier transform of the convolutions in the Volterra series reduces the order of the computation. For example, the first convolution integral in (2), representing the linear portion of the system, is an $O(N^2)$ computation, while its Fourier transform results in a simple multiplication, which is $O(N)$, as shown in (4).

$$y_1(t) = \int_{-\infty}^{\infty} h_1(\tau) x(t-\tau) d\tau \tag{4a}$$

$$Y_1(f) = H_1(f) X(f). \tag{4b}$$

The three-dimensional Fourier transform of the third-order Volterra operator enables a similar frequency-domain computation [10], shown in (5a)–(5b) at the bottom of the page.

Equation (5b) can be used for intermodulation distortion (IM_3) calculations, which involve only two tones. In this case, $X(f)$ is nonzero at only four frequency values, two on the positive axis (f_1 and f_2) and two on the negative axis ($-f_2$ and $-f_1$). Therefore, the double integral in (5b) reduces to a few multiplications and additions. This has allowed designers in the past to perform hand calculations based on the Volterra series [11]. There are well-known procedures, such as the Bussgang method, allowing the computation of a symmetric $H_3(\omega_1, \omega_2, \omega_3)$ based on the SPICE parameters [12]. This is the method utilized for calculating the Volterra kernels in this treatment.

As explained above, $X(f)$ is no longer a deterministic signal when an analysis of the spectral regrowth using modulated signals is desired. Even though frequency-domain computation reduces the required number of samples and the order of the computation compared to the time-domain approach, implementing (5b) directly would still result in an $O(N^3)$ algorithm. The resulting computation times are too long to be practical, which sometimes lead researchers to assume the Volterra kernels to be constant, reducing the Volterra series to

$$y(t) = \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) d\tau_1 d\tau_2 + \cdots$$
$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(\tau_1, \tau_2, \ldots, \tau_n) x(t-\tau_1) x(t-\tau_2)\cdots x(t-\tau_n) d\tau_1 d\tau_2 \cdots d\tau_n + \cdots \tag{2}$$

$$y_3(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) x(t-\tau_1) x(t-\tau_2) x(t-\tau_3)\, d\tau_1\, d\tau_2\, d\tau_3 \tag{5a}$$

$$Y_3(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_3(\alpha, \beta-\alpha, f-\beta) X(\alpha) X(\beta-\alpha) X(f-\beta)\, d\alpha\, d\beta. \tag{5b}$$

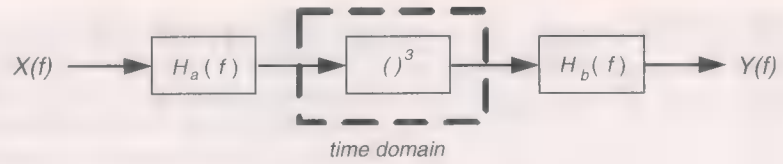


Fig. 1. Pure third-order subsystem.

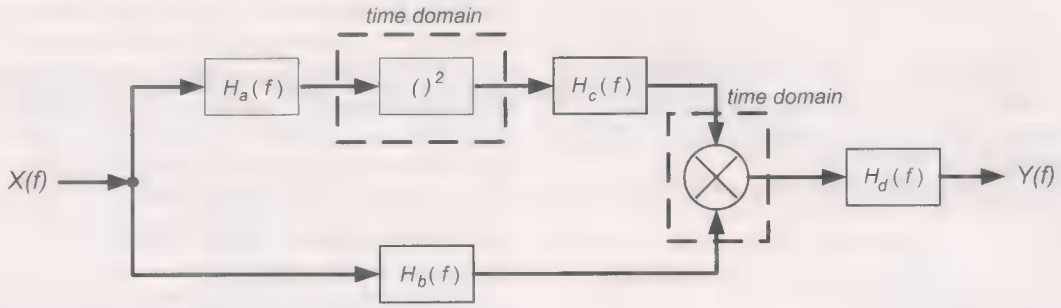


Fig. 2. Second-order interaction subsystem.

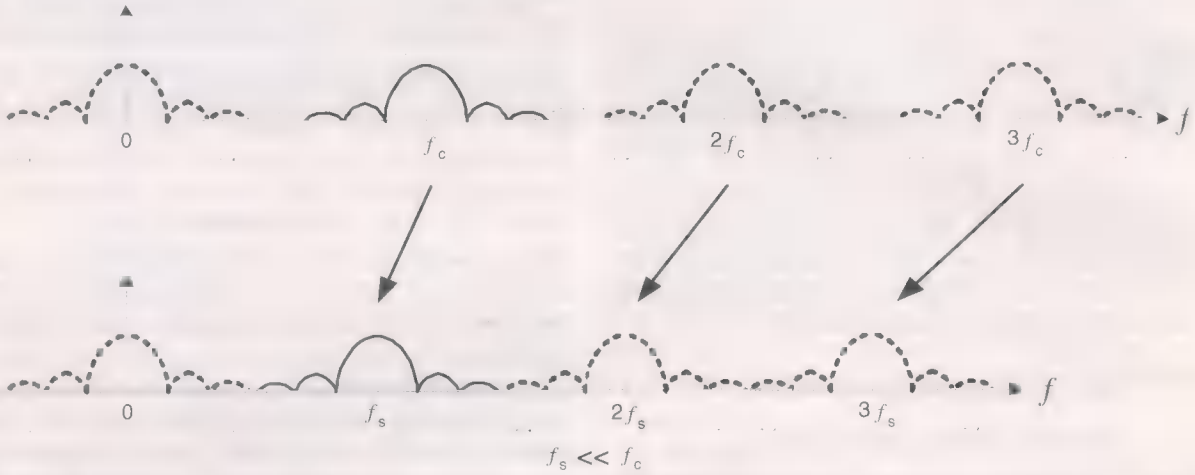


Fig. 3. Compressed spectrum.

a Taylor series expansion with complex coefficients [13]. However, this ignores memory effects arising from second-order interaction terms which represent the mixing of linear signals with second-order distortion products if there is feedback or nonlinearities are cascaded in the circuit.

B. Decomposition of the Volterra Kernels

In the past, the long computation times have prevented the Volterra series from being utilized in a distortion analysis involving more than a few input tones. In order to dramatically reduce the amount of time it takes for the computations to be completed, a decomposition of the Volterra kernels is proposed in this treatment. A closer examination shows that $H_3(\omega_1, \omega_2, \omega_3)$ for circuits can be broken down into parallel combinations of models resembling the ones shown in Figs. 1 and 2 without using any approximations [9]. The former figure shows the pure third-order subsystem, while the latter one represents the second-order interaction. This decomposition is still an exact representation of the original Volterra system as shown in the Appendix , because the only source of memory in the PA circuit is the frequency dependence of the impedance of inductors or capacitors (whether their values are constant or voltage-dependent.)

This decomposition allows the representation of the third-order nonlinear system with memory by a combination of some linear blocks with memory and nonlinear blocks without memory. The computations involving the linear blocks represented by filters can easily be done in the frequency-domain according to (4b). As the nonlinear blocks lack memory, cubing, squaring or multiplication of the signals can be done in the time-domain at a simulation carrier frequency f_s much lower than the actual carrier frequency f_o. This allows the compression of the spectrum as shown in Fig. 3. The combination of time and frequency-domain calculations allows the response of the circuit to be represented by a closed-form solution, instead of requiring numerical solutions of differential equations described in [14].

The compressed spectrum requires about 8 times more samples than the baseband equivalent version, but the order of the computation is reduced to $O(N \log N)$ limited by the inverse-FFT and FFT calculations required before and after the nonlinear blocks. Therefore, all of the computations can

be completed in a very short time frame using a tool such as Matlab [15]. It is possible to reduce the processing time further by using a more direct programming language such as C if ease of implementation is not sought.

C. Device Modeling

The new computational approach described above was applied to the design of PAs using bipolar transistors as the active element. The nonlinear bipolar transistor model used in the analysis is shown in Fig. 4. This model is applicable to Si and SiGe BJTs, as well as GaAs HBTs. The output impedance seen by a PA is generally small compared to $r_o = I_c/V_A$, where V_A is the Early voltage, so r_o is neglected in the model. The parasitic capacitance between collector and substrate can be assumed small in a silicon-on-insulator (SOI) process or can be assumed constant and combined with the package parasitics and output matching network in more conventional processes.

The parasitic collector resistance r_c and emitter resistance r_e are assumed constant. Although the value of r_b is known to change at high base-current levels, its nonlinearity is usually negligible compared to the other nonlinear elements in the transistor [12], so r_b is assumed constant as well. The transconductance $g_m = I_c/V_T$ is nonlinear due to the exponential relationship between I_c and V_{BE}. The base-emitter resistance $r_\pi = \beta/g_m$ is also nonlinear, because of g_m. The variation of β due to changes in the base-current is assumed negligible compared to the exponential behavior of g_m, so β is assumed constant.

Linear PA design requires that the power transistors are prevented from going into the saturation region. Therefore, it is assumed that the collector-base junction is never forward biased for the analysis. Thus, the collector-base capacitor C_μ in Fig. 4 is composed of some constant parasitic capacitance and the collector-base junction depletion capacitance. The latter capacitance is the cause of nonlinearity in C_μ and the analysis shows that its contribution to distortion is considerable due to the large signal swings across the collector-base junction during the PA operation.

The base-emitter capacitance C_π in Fig. 4 is given by

$$C_\pi = C_b + C_{je} = g_m \tau_F + C_{je} \tag{6}$$

where C_{je} is the base-emitter depletion capacitance, C_b is the diffusion capacitance and τ_F is the forward transit time. Although C_{je} is usually neglected or assumed constant, its value can become comparable to C_b and its nonlinearity can be significant when V_{BE} swing is large. Therefore, it is necessary to model the nonlinearity of C_{je} for an accurate analysis.

D. Modeling the Variations in Forward Transit Time

The diffusion capacitance C_b also varies with V_{BE} because of the variation in the transconductance g_m and forward transit time τ_F. The latter is usually assumed constant, but its value starts to increase and cause additional distortion at high collector current densities. In order to increase the accuracy of the analysis, the series based model outlined below has been developed by the authors to take this variation into account.

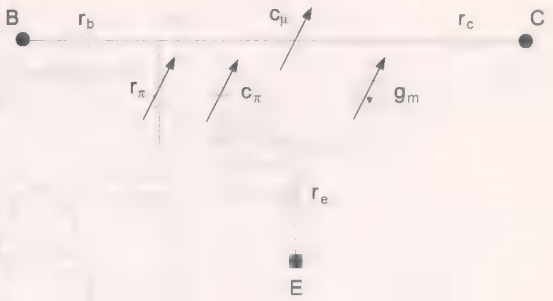


Fig. 4. Nonlinear bipolar transistor model.

The forward transit time τ_F has four components

$$\tau_F = \tau_E + \tau_{BE} + \tau_B + \tau_{BC} \tag{7}$$

where τ_E is the emitter transit time, τ_{BE} is the base-emitter depletion region transit time, τ_B is the base transit time, and τ_{BC} is the base-collector depletion region transit time [16].

The first component of τ_F affected by the increasing current density is usually τ_{BC}. When current is flowing through an npn transistor, the injected electrons are added to the negatively charged depletion region on the base side and subtracted from the positively charged depletion region on the collector side [17]. For a constant base-collector voltage, this requires the depletion region of the collector to be wider and τ_{BC} to be larger. The depletion region of the base becomes shorter as well, but the effect is much less pronounced as the doping density in the base is several orders of magnitude higher than in the collector. If the current density is increased further, other mechanisms, such as base widening or Kirk effect, will be observed [16], [18], [19]. However, at this point, the rise in τ_F is quite rapid and this operating region is not suitable for a linear PA. Therefore, only the variation in τ_{BC} is modeled for this work.

If the same basic assumptions outlined in [17] are followed, τ_{BC} can be calculated as

$$\tau_{BC} = \frac{q x_{d0} N_c}{J_c} \left(\sqrt{\frac{1}{1 - \frac{J_c}{q N_c \vartheta_{sat}}}} - 1 \right) \tag{8}$$

where q is the electron charge, x_{d0} is the width of the depletion region without current, N_c is the collector dopant density, J_c is the collector current density and ϑ_{sat} is the scattering-limited velocity of carriers.

In this case, τ_F can be expanded as a Taylor series

$$\begin{aligned} &\tau_F(V_{BE} + v_{be}) \\ &= \tau_F(V_{BE}) + \tau'_{BC}(V_{BE}) v_{be} + \frac{\tau''_{BC}(V_{BE})}{2} v_{be}^2 \\ &\quad + \frac{\tau'''_{BC}(V_{BE})}{6} v_{be}^3 + \cdots \\ &= \tau_F(V_{BE}) + K_{\tau_{F1}} v_{be} + K_{\tau_{F2}} v_{be}^2 + K_{\tau_{F3}} v_{be}^3 + \cdots . \end{aligned} \tag{9}$$

Thus, the current in C_b is given by

$$i_b = \frac{dQ_b}{dt} = \frac{d}{dt}(\tau_F I_c) = \tau_F \frac{dI_c}{dt} + I_c \frac{d\tau_F}{dt}$$

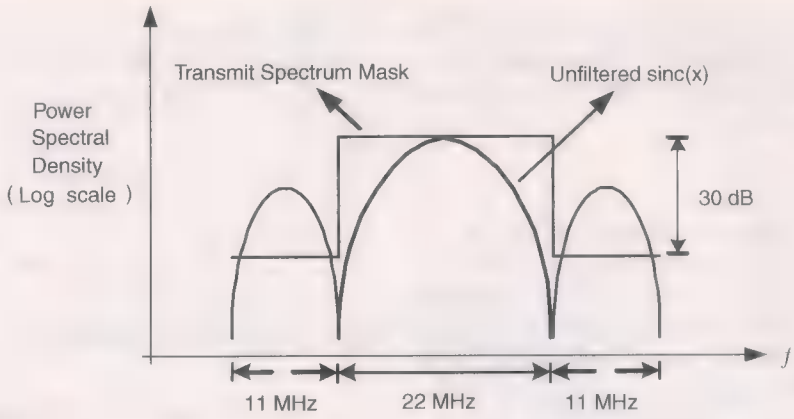


Fig. 5. IEEE802.11b transmit spectrum mask.

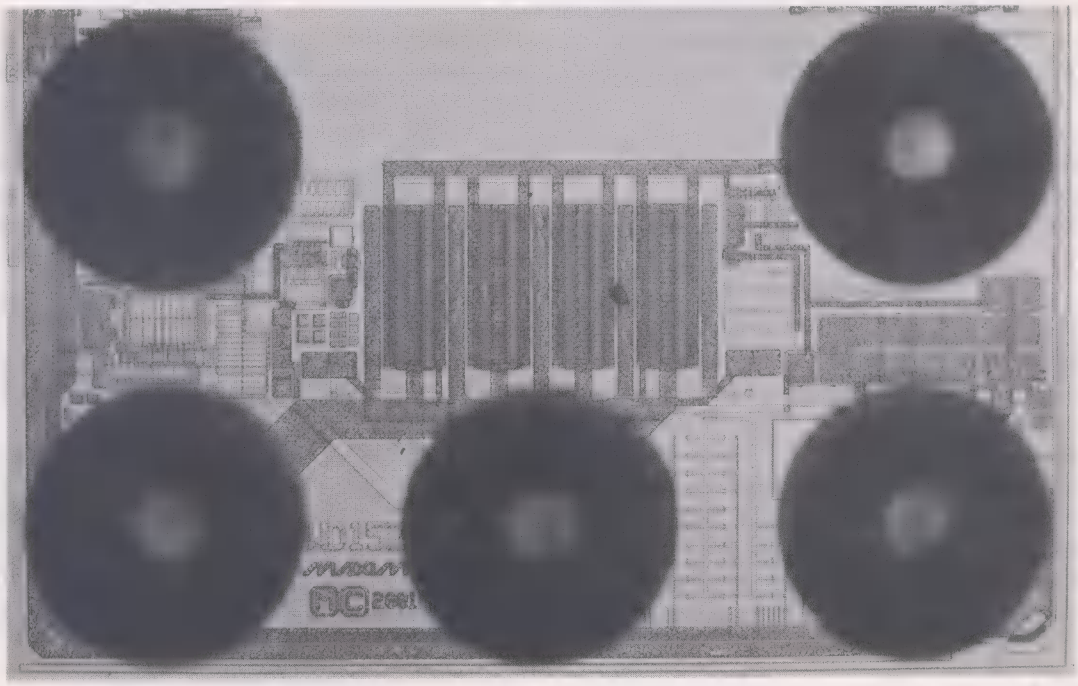


Fig. 6. Die photo with the power transistor in the center and solder bumps around it.

$$= \tau_F \left(g_m \frac{dv_{be}}{dt} + K_{g_{m2}} \frac{dv_{be}^2}{dt} + K_{g_{m3}} \frac{dv_{be}^3}{dt} + \cdots \right) + I_c \left(K_{\tau_{F1}} \frac{dv_{be}}{dt} + K_{\tau_{F2}} \frac{dv_{be}^2}{dt} + K_{\tau_{F3}} \frac{dv_{be}^3}{dt} + \cdots \right). \quad (10)$$

The first part of (10) is the usual equation which governs the Volterra series representing the nonlinearity of the diffusion capacitance when τ_F is assumed constant. The second part is the series proposed for modeling the variations of τ_{BC}. Equation (10) can be further simplified as

$$i_b = (g_m \tau_F + I_c K_{\tau_{F1}}) \frac{dv_{be}}{dt} + (K_{g_{m2}} \tau_F + I_c K_{\tau_{F2}}) \frac{dv_{be}^2}{dt} + (K_{g_{m3}} \tau_F + I_c K_{\tau_{F3}}) \frac{dv_{be}^3}{dt} + \cdots \quad (11)$$

so that τ_F variation model does not increase the computational complexity, once the extra K_{τ_F} coefficients are calculated.

IV. Implementation

In order to compare the results of the analysis with measurements, a number of single-stage PAs have been designed for the IEEE802.11b wireless LAN standard operating at 2.4 GHz in the ISM band. The standard specifies the maximum output power level to be 20 dBm at the antenna, but the PAs have been designed to supply 24 dBm in order to accommodate the losses through the passive elements before the signal reaches the antenna. The transmit spectrum mask specifications require the spectral products in the adjacent sidelobe to be 30 dB below the main lobe, as shown in Fig. 5.

The PAs have been designed using SiGe bipolar transistors and flip-chip packaging. Measurements have been taken using different values for input and output matching elements, bias current, supply voltage, as well as different number of transistors, which can be changed by means of a laser cutter. The die photo is shown in Fig. 6.

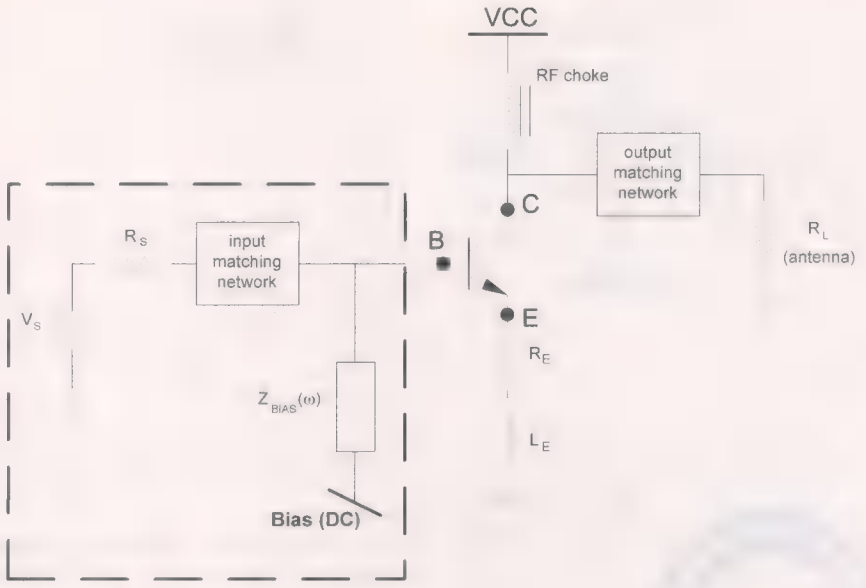


Fig. 7. Simplified schematic of the power amplifier.

A. Details of the Circuit

The simplified schematic of the common emitter PA is shown in Fig. 7. In order to reduce the number of nodes, the Norton equivalent of the signal source, input matching and local-bias circuit, shown in the dashed box, has been used in the analysis.

The off-chip choke between the collector and V_{CC} passes the bias current, but has high impedance at RF frequencies so that almost all of the signal flows into the antenna through the output matching network. R_E includes the parasitic emitter resistance, as well as the emitter degeneration resistance. The total value of R_E is adjusted so that there is about 50 mV dc voltage drop across it, in order to prevent thermal runaway. L_E is composed of the on-chip wiring inductance, package and board parasitics. The on-chip part is estimated by using Greenhouse formulas [20] and the off-chip parts are calculated based on a 3-D EM simulator. Both R_E and L_E improve linearity through series feedback, but an inductor is preferred, as it does not limit voltage headroom as a resistor does. Inductive degeneration also increases the real part of the input impedance so that the input (or interstage) matching network can be designed with a lower quality factor (Q).

The series feedback used to improve the linearity of the PA usually causes the third-order coefficient of the system to have a sign opposite of the first-order one. Thus, gain compression occurs at high power levels. One way to alleviate the gain compression problem is to allow the PA bias current to increase at high power levels through some modifications of the biasing circuitry. The conventional biasing circuitry for a common emitter amplifier is usually based on a current mirror with a current helper as shown in Fig. 8 [21]. The ratio of the resistors tied to the base and emitter of Q_1 and Q_2 must be adjusted to make sure the voltage drop across them and, in turn, the voltage drop across the base-emitter junctions are the same. The base resistor and capacitor C_F act as a filter to attenuate the input signal before reaching the bias circuit.

A large input signal swing increases the average value of the collector current over the quiescent value, due to the exponential

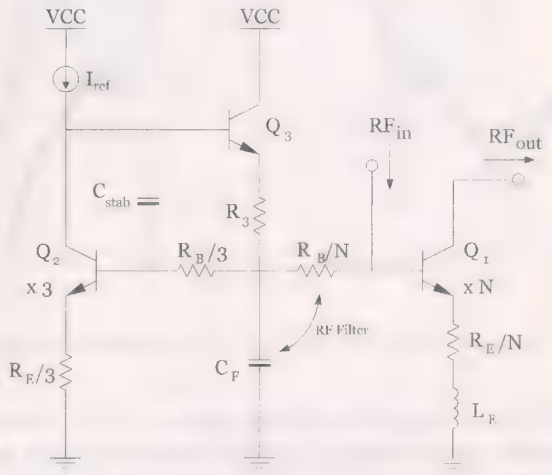


Fig. 8. Conventional local biasing circuit.

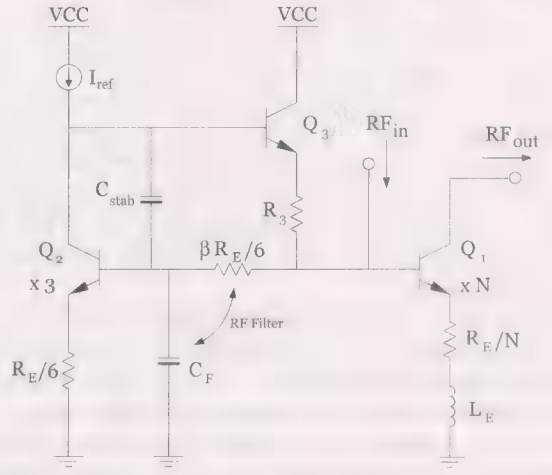


Fig. 9. Local biasing circuit without base resistor.

nature of the bipolar transistor. However, this increases the base current and causes a bigger voltage drop across the base resistor, reducing the base-emitter voltage of the power transistor Q_1 and

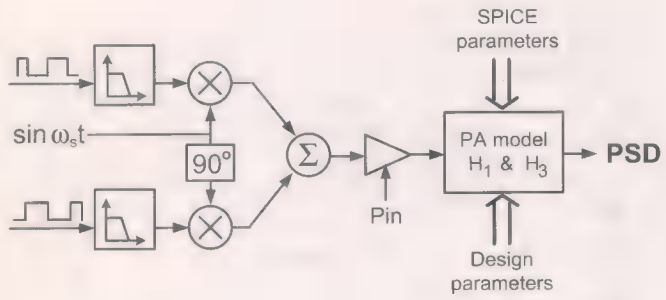


Fig. 10. Simulation method.

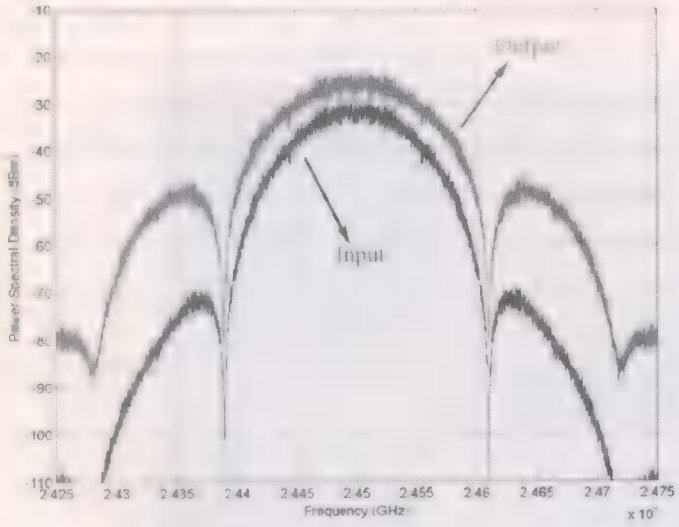


Fig. 11. Input and output power spectral density calculated by simulations.

limiting the increase in the average collector current. Therefore, the biasing circuit shown in Fig. 9 is preferred, as the lack of a resistor at the base of the power transistor allows a bigger increase in average collector current at high power levels. A further advantage is that the lower impedance seen by the base of Q_1 increases the device breakdown voltage. The values of the emitter and base resistors of Q_2 need to be adjusted so that the voltage drop across the base-emitter junctions of both transistors are equal to each other. The current mirror ratio depends on the actual value of β, but dc simulations has shown that the change in the collector current of Q_1 due to the process and temperature variations is still small. C_F also needs to be moved toward the base of Q_2, so that the signal is not attenuated before reaching the power transistor.

Analysis shows that the amount of output voltage swing at a given output power level is quite sensitive to the imaginary part of the impedance at the collector. A small positive imaginary part at that node can easily increase the voltage swing across base-collector junction and push the transistor into saturation, even if the output return loss S_{22} stays acceptable. Therefore a transmission line, instead of a discrete inductor, is used in the output matching network for more precise tuning. The typical length of the transmission line required for a PA in a 50-Ω system is usually short enough to be realized on compact circuit boards.

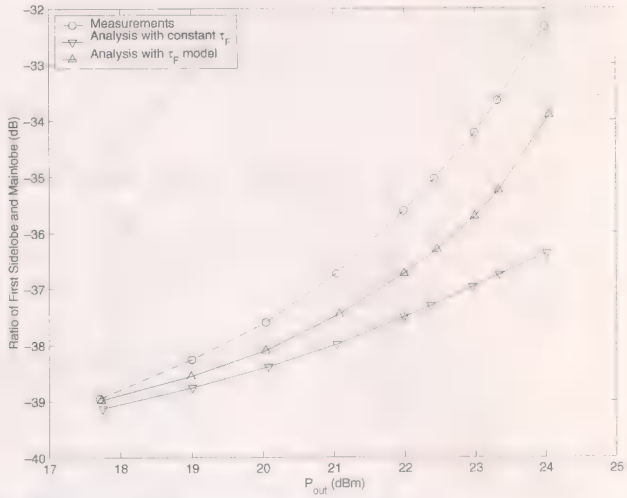


Fig. 12. $m = 78$, $V_{CC} = 3.3$ V, $I_{cbias} = 196$ mA, $f_c = 2.45$ GHz.

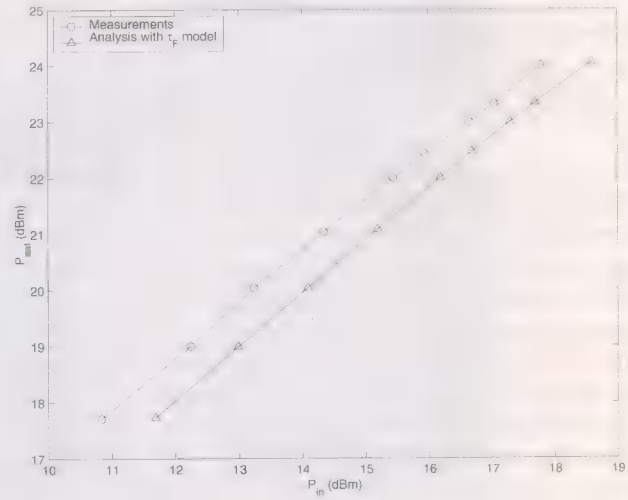


Fig. 13. $m = 78$, $V_{CC} = 3.3$ V, $I_{cbias} = 196$ mA, $f_c = 2.45$ GHz, 802.11b modulation.

V. Results

The simulation method using the Volterra-series-based power-amplifier model and the decomposition of the Volterra kernels has been implemented in Matlab. Baseband I and$Q$ signals are generated from oversampled 1024-bit-long random data streams filtered according to the specified modulation scheme. The baseband signals are then upconverted to the simulation carrier frequency to generate the input waveform, as shown in Fig. 10. The input signal amplitude is adjusted according to the desired input power level and fed into the PA model, which includes the SPICE coefficients of the npn transistors and the design parameters, to generate the output waveform [22]. The power spectral densities of the input and output waveforms shown in Fig. 11 are generated by this method. The amplitude of the adjacent sidelobes relative to the mainlobe is much larger in the output than in the input due to spectral regrowth.

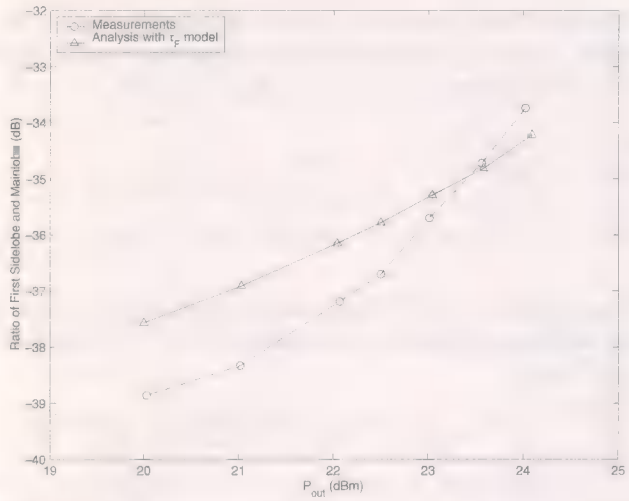


Fig. 14. $m = 78$, $V_{CC} = 3.5$ V, $I_{cbias} = 176$ mA, $f_c = 2.0$ GHz.

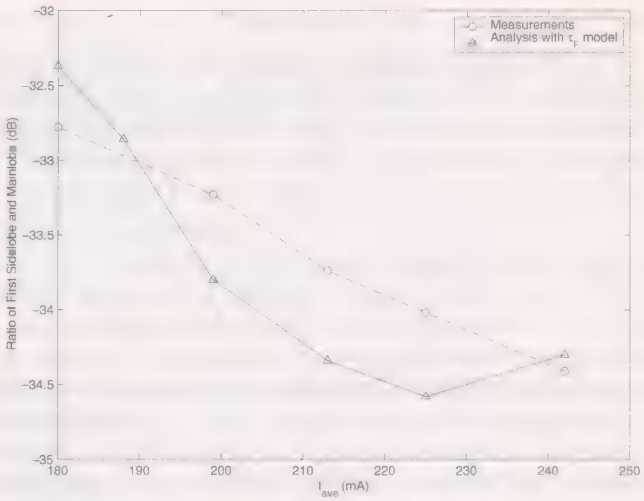


Fig. 15. $P_{out} = 24$ dBm, $m = 78$, $V_{CC} = 3.5$ V, $f_c = 2.0$ GHz.

The measured ratio of the adjacent sidelobe and mainlobe versus output power level are compared with numerical predictions in Fig. 12. An IEEE802.11b modulated waveform at a carrier frequency of 2.45 GHz is applied to a power amplifier which consists of 78 output transistors in parallel. The PA is operated at a supply voltage of 3.3 V and a bias current of 196 mA. At high power levels the current density becomes large and an assumption of constant τ_F model results in underestimation of the spectral regrowth. The analysis including the τ_F predicts the measured sidelobe growth to within 1.6 dB or better accuracy, while only requiring minimal computation time. The predicted and measured gain with IEEE802.11b modulated waveform at different power levels are shown in Fig. 13.

The results of the analysis show that the increase in the base-collector depletion region transit time τ_{BC} at high collector current densities can easily become the dominant source of nonlinearity in a power amplifier. This problem can be alleviated by increasing the total emitter area of the PA at the expense of increased parasitics and lower gain. The analysis predicts that using 104 parallel output transistors would reduce the contribution of τ_{BC} variation to negligible levels and improve spectral regrowth, which agrees with the measurements. The spectral regrowth can also be improved by some modifications to the power transistor, such as increasing the collector dopant density or placing the highly doped buried layer closer to the collector-base junction, but attention must be paid by the device designers to make sure the device breakdown voltage does not become too low.

A number of similar measurements have been taken from another PA with 78 parallel output transistors. The results of the measurements and analysis for this case can be seen in Fig. 14. The operating frequency in this case is 2 GHz, the quiescent current is 176 mA and the supply voltage is 3.5 V. The predicted spectral regrowth again differs by less than 1.5 dB compared to the measurements. Fig. 15 shows the results for the same PA supplying 24 dBm output at different average current levels. The measurements and predictions agree to within 0.6 dB. The trend

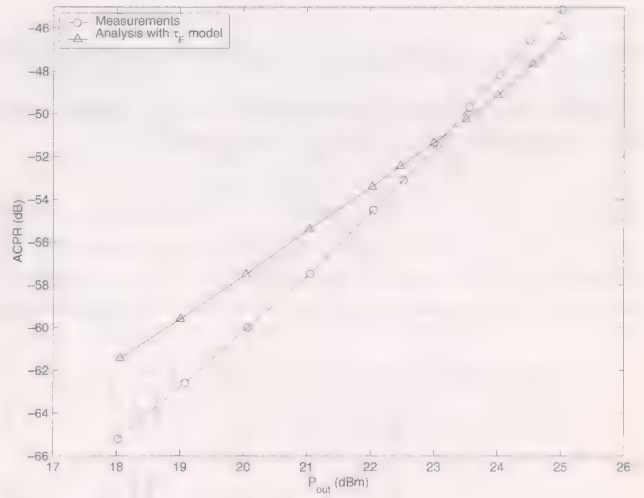


Fig. 16. IS-95 modulation, $m = 78$, $V_{CC} = 3.5$ V, $I_{cbias} = 176$ mA, $f_c = 2.0$ GHz.

difference at very high currents is due to the onset of Kirk effect and consequent modeling inaccuracies, which result in measured distortion to be somewhat below the predicted value.

The agreement between the measurements and analysis is similar when the same PA is operated with IS-95 waveforms at the same carrier frequency as shown in Fig. 16. The expected ACPR differs by less than 2 dB compared to the measurements.

This analysis has also been used to compare resistive and inductive degeneration. It shows that in addition to providing more headroom and higher input impedance, inductive degeneration improves spectral regrowth considerably as well. This result agrees with a similar prediction for IM_3 improvement [23]. Another linearity improvement method is using a low-frequency-trap network [24]. However, the resistor used for preventing thermal runaway makes the improvement in spectral regrowth to be very small, in agreement with the analysis done for the LNA in [24]. It should also be pointed out that even if resistive degeneration is not used, the improvement in spectral regrowth through the use of a low-frequency-trap network is reduced due to τ_F variation.

VI. Conclusion

A novel method of analyzing spectral regrowth based on the Volterra series and basic SPICE parameters has been developed. The proposed decomposition of the Volterra kernels into simpler subsystems have dramatically reduced the computation times. A series based model has also been developed to represent the increase in the forward transit time of bipolar transistors at high collector current densities. A number of single stage SiGe power amplifiers have been designed, fabricated and tested to validate the analysis. The computations based on this method provides a good insight into the relationship between spectral regrowth and the physical mechanisms in bipolar transistors. This can help circuit designers better understand the effect of design parameters on spectral regrowth, as well as the design trade-off between efficiency and linearity. In addition, it can help device designers optimize power transistors by better identifying the transistor components contributing most to distortion.

Appendix I
Neglecting Even-Order Terms in Volterra Series

The input $x(t)$ to the system described by the Volterra series can be represented by an inverse Fourier transform

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df. \tag{12}$$

The nth order frequency domain Volterra kernel of that system can be expressed as the n-dimensional Fourier transform of the time-domain kernel, $h_n(\tau_1, \ldots, \tau_n)$

$$H_n(f_1, \ldots, f_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) e^{-j2\pi(f_1\tau_1 + \cdots + f_n\tau_n)} d\tau_1 \cdots d\tau_n. \tag{13}$$

The nth term of the Volterra series can then be rewritten as (14), shown at the bottom of the page, if $x(t-\tau_i)$ is replaced by (12) and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n)\, e^{-j2\pi \sum_{i=1}^{n} \rho_i \tau_i}\, d\tau_1 \cdots d\tau_n$$

by (13).

Variable ρ_n can be replaced by defining $f = \rho_1 + \cdots + \rho_{n-1} + \rho_n$. Please note that $df = d\rho_n$ and $\rho_n = f - \sum_{i=1}^{n-1} \rho_i$. Thus, (14) becomes (15), also shown at the bottom of the page.

If (15) is compared to the inverse Fourier transform equation

$$y_n(t) = \int_{-\infty}^{\infty} Y_n(f) e^{j2\pi ft} df \tag{16}$$

the Fourier transform of the nth term of the Volterra series can be calculated as (17), also shown at the bottom of the page.

$$\begin{aligned}
y_n(t) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \left[\prod_{i=1}^{n} x(t-\tau_i) \right] d\tau_1 \cdots d\tau_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \left[\prod_{i=1}^{n} \int_{-\infty}^{\infty} X(\rho_i) e^{j2\pi\rho_i(t-\tau_i)}\, d\rho_i \right] d\tau_1 \cdots d\tau_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n)\, e^{-j2\pi \sum_{i=1}^{n} \rho_i \tau_i}\, d\tau_1 \cdots d\tau_n \right] \left[\prod_{i=1}^{n} X(\rho_i) \right] e^{j2\pi \sum_{i=1}^{n} \rho_i t}\, d\rho_1 \cdots d\rho_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(\rho_1, \ldots, \rho_n) \left[\prod_{i=1}^{n} X(\rho_i) \right] e^{j2\pi \sum_{i=1}^{n} \rho_i t}\, d\rho_1 \cdots d\rho_n
\end{aligned} \tag{14}$$

$$y_n(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n\left(\rho_1, \ldots, \rho_{n-1}, f - \sum_{i=1}^{n-1} \rho_i\right) X(\rho_1) \cdots X(\rho_{n-1}) X\left(f - \sum_{i=1}^{n-1} \rho_i\right) e^{j2\pi ft}\, d\rho_1 \cdots d\rho_{n-1} df \tag{15}$$

$$Y_n(f) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n\left(f_1, \ldots, f_{n-1}, f - \sum_{i=1}^{n-1} f_i\right) X(f_1) \cdots X(f_{n-1}) X\left(f - \sum_{i=1}^{n-1} f_i\right) df_1 \cdots df_{n-1} \tag{17}$$

Let us assume that a narrowband input signal is applied to this nonlinear system, such that the carrier frequency f_o is much greater than the bandwidth of the signal Δf. Hence, $X(f_i)$ is nonzero only for

$$f_i \in \left[\left(-f_o - \frac{\Delta f}{2}, -f_o + \frac{\Delta f}{2}\right) \cup \left(f_o - \frac{\Delta f}{2}, f_o + \frac{\Delta f}{2}\right)\right]$$

where $\Delta f \ll f_o$. Thus, the last term of the integral in (17) is nonzero only if

$$f - \sum_{i=1}^{n-1} f_i \in \left[\left(-f_o - \frac{\Delta f}{2}, -f_o + \frac{\Delta f}{2}\right) \cup \left(f_o - \frac{\Delta f}{2}, f_o + \frac{\Delta f}{2}\right)\right]. \tag{18}$$

Therefore, $Y_n(f)$ is nonzero around f_o only if $\sum_{i=1}^{n-1} f_i$ is about zero. This requires $(n-1)/2$ of f_i terms to be in $(-f_o - (\Delta f/2), -f_o + (\Delta f/2))$ and the remaining $(n-1)/2$ terms to be in $(f_o - (\Delta f/2), f_o + (\Delta f/2))$. In order for $(n-1)/2$ to be an integer value, n has to be odd.

Appendix II
Third-Order Volterra Subsystems

A. Second-Order Interaction Subsystem

Analyzing the second-order interaction subsystem shown in Fig. 2 is easier when the upper path is considered first. If the input and output of the squarer are called $U(f)$ and $V(f)$, respectively, the following relationships can be derived:

$$U(f) = H_a(f)X(f) \tag{19}$$

and

$$\begin{aligned} V(f) = U(f) * U(f) &= \int_{-\infty}^{\infty} U(\rho)U(f-\rho)\,d\rho \\ &= \int_{-\infty}^{\infty} H_a(\rho)X(\rho)H_a(f-\rho)X(f-\rho)\,d\rho \end{aligned} \tag{20}$$

as a multiplication in the time domain is equivalent to a convolution in the frequency domain. Applying this property once again, the output of the multiplier $Z(f)$ can be calculated as shown in (21) at the bottom of the page. As $Z(f)$ can also be defined as

$$\begin{aligned} &Z(f) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_Z(\alpha, \beta-\alpha, f-\beta)X(\alpha)X(\beta-\alpha)X(f-\beta)\,d\alpha\,d\beta \end{aligned} \tag{22}$$

some simple change of variables allow $H_Z(f_1, f_2, f_3)$ to be expressed as

$$H_Z(f_1, f_2, f_3) = H_c(f_1 + f_2)H_b(f_3)H_a(f_1)H_a(f_2). \tag{23}$$

Unfortunately, this kernel is not symmetric. In order to get a kernel which does not depend on the exact order of the variables (f_1, f_2, f_3), (21) needs to be rewritten as (24), also shown at the bottom of the page. The variables ρ and F in the first double integral of (24) can easily be replaced by α and β, respectively. The variables ρ and $f - F$ in the second one can then be replaced by α and $\beta - \alpha$, respectively. Similarly, the variables $f - \rho$ and $f - F$ in the last double integral can be replaced by β and α, respectively, making $d\beta = -d\rho$ and $d\alpha = -dF$. After making the appropriate changes in the limits of the integrals, (24) can be represented by (25), shown at the top of the next page, and (23) becomes

$$\begin{aligned} H_Z(f_1, f_2, f_3) = \frac{1}{3}\big(&H_c(f_1+f_2)H_b(f_3)H_a(f_1)H_a(f_2) \\ &+ H_c(f_1+f_3)H_b(f_2)H_a(f_1)H_a(f_3) \\ &+ H_c(f_2+f_3)H_b(f_1)H_a(f_2)H_a(f_3)\big). \end{aligned} \tag{26}$$

$$\begin{aligned} Z(f) &= \big(H_b(f)X(f)\big) * \big(H_c(f)V(F)\big) \\ &= \int_{-\infty}^{\infty} H_b(f-F)X(f-F)H_c(F)V(F)\,dF \\ &= \int_{-\infty}^{\infty} H_b(f-F)X(f-F)H_c(F)\left[\int_{-\infty}^{\infty} H_a(\rho)X(\rho)H_a(F-\rho)X(F-\rho)\,d\rho\right] dF \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F)\,d\rho\,dF \end{aligned} \tag{21}$$

$$\begin{aligned} Z(f) &= \frac{1}{3}\big(Z(f) + Z(f) + Z(f)\big) \\ &= \frac{1}{3}\bigg(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F)\,d\rho\,dF \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F)\,d\rho\,dF \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F)\,d\rho\,dF\bigg) \end{aligned} \tag{24}$$

$$Z(f) = \frac{1}{3}\Bigg(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(\beta)H_b(f-\beta)H_a(\alpha)H_a(\beta-\alpha)X(\alpha)X(\beta-\alpha)X(f-\beta)d\rho\, dF + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(f-\beta+\alpha)H_b(\beta-\alpha)H_a(\alpha)H_a(f-\beta)X(\alpha)X(f-\beta)X(\beta-\alpha)\, d\rho\, dF + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_c(f-\alpha)H_b(\alpha)H_a(f-\beta)H_a(\beta-\alpha)X(f-\beta)X(\beta-\alpha)X(\alpha)\, d\rho\, dF \Bigg) \tag{25}$$

$$\begin{aligned} Y(f) &= H_b(f)V(f) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_b(f)U(\alpha)U(\beta-\alpha)U(f-\beta)\, d\alpha\, d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_b(f)H_a(\alpha)H_a(\beta-\alpha)H_a(f-\beta)X(\alpha)X(\beta-\alpha)X(f-\beta)\, d\alpha\, d\beta \end{aligned} \tag{30}$$

As the output of the second-order interaction subsystem $Y(f)$ is

$$Y(f) = H_d(f)Z(f) \tag{27}$$

$H_3(f_1, f_2, f_3)$ of this subsystem can be calculated as

$$\begin{aligned} &H_3(f_1, f_2, f_3) \\ &= \frac{H_d(f_1+f_2+f_3)}{3}\big(H_c(f_1+f_2)H_b(f_3)H_a(f_1)H_a(f_2) \\ &\quad + H_c(f_1+f_3)H_b(f_2)H_a(f_1)H_a(f_3) \\ &\quad + H_c(f_2+f_3)H_b(f_1)H_a(f_2)H_a(f_3)\big). \end{aligned} \tag{28}$$

B. Pure Third-Order Subsystem

A similar but simpler analysis can be performed for the pure third-order subsystem shown in Fig. 1. If the input and output of the cuber are called $U(f)$ and $V(f)$, respectively, $V(f)$ can be expressed as

$$\begin{aligned} V(f) &= U(f) * U(f) * U(f) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(\alpha)U(\beta-\alpha)U(f-\beta)d\alpha\, d\beta \end{aligned} \tag{29}$$

by using the property of the frequency convolution. Thus, $Y(f)$ is given by (30), shown at the top of the page. As $Y_3(f)$ is defined as

$$\begin{aligned} &Y_3(f) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_3(\alpha, \beta-\alpha, f-\beta)\, X(\alpha)\, X(\beta-\alpha)X(f-\beta)\, d\alpha\, d\beta \end{aligned} \tag{31}$$

some simple change of variables allow $H_3(f_1, f_2, f_3)$ of this subsystem to be represented by

$$H_3(f_1, f_2, f_3) = H_b(f_1+f_2+f_3)H_a(f_1)H_a(f_2)H_a(f_3). \tag{32}$$

When the third-order Volterra kernels encountered during an analysis of a circuit [12] is compared to (28) and (32), it can easily be shown that the original Volterra system can be decomposed into parallel combinations of the subsystems resembling the ones shown in Figs. 1 and 2 without any approximations.

ACKNOWLEDGMENT

The authors would like to thank Maxim Integrated Products for the assistance with manufacturing and testing the power amplifiers. The authors are also grateful to J. King for valuable technical discussions.

REFERENCES

[1] S. W. Chen, W. Panton, and R. Gilmore, "Effects of nonlinear distortion on CDMA communications systems," *IEEE Trans. Microwave Theory Tech.*, vol. 44, no. 12, pp. 2743–2749, Dec. 1996.

[2] S. Pinsky, "A method for computing adjacent-channel spectral energy in cellular power amplifiers," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1998, pp. 1595–1598.

[3] W. Struble, F. McGrath, K. Harrington, and P. Nagle, "Understanding linearity in wireless communication amplifiers," *IEEE J. Solid-State Circuits*, vol. 32, no. 9, pp. 1310–1318, Sep. 1997.

[4] A. Leke and J. S. Kenney, "Behavioral modeling of narrowband microwave power amplifiers with applications in simulating spectral regrowth," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1996, pp. 1385–1387.

[5] Q. Wu, M. Testa, and R. Larkin, "Linear RF power amplifier design for CDMA signals," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1996, pp. 851–854.

[6] K. G. Gard, H. M. Gutierrez, and M. B. Steer, "Characterization of spectral regrowth in microwave amplifiers based on the nonlinear transformation of a complex Gaussian process," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 7, pp. 1059–1069, Jul. 1999.

[7] G. T. Zhou, "Analysis of spectral regrowth of weakly nonlinear power amplifiers," *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 357–359, Nov. 2000.

[8] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proc. IEEE*, vol. 69, no. 12, pp. 1557–1573, Dec. 1981.

[9] B. Baytekin, "Analysis and design of monolithic radio frequency linear power amplifiers," Ph.D. dissertation, University of California, Berkeley, 2004.

[10] J. S. Bendat, *Nonlinear System Techniques and Applications*. New York: Wiley, 1998.

[11] S. Narayanan, "Intermodulation distortion of cascaded transistors," *IEEE J. Solid-State Circuits*, vol. 4, no. 3, pp. 97–106, Jun. 1969.

[12] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*. Boston, MA: Kluwer, 1998.
[13] S. A. Maas, "Volterra analysis of spectral regrowth," *IEEE Microwave Guided Wave Lett.*, vol. 7, no. 7, pp. 192–193, Jul. 1997.
[14] V. Borich, J. East, and G. Haddad, "The method of envelope currents for rapid simulation of weakly nonlinear communication circuits," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1999, pp. 981–984.
[15] *Matlab 6.1*, The Mathworks, Inc., Natick, MA.
[16] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[17] R. G. Meyer and R. S. Muller, "Charge control analysis of the collector-base space-charge region contribution to bipolar-transistor time constant τ_T," *IEEE Trans. Electron Devices*, vol. 34, no. 2, pp. 450–452, Feb. 1987.
[18] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed. New York: Wiley, 1986.
[19] D. J. Roulston, *Bipolar Semiconductor Devices*. New York: McGraw-Hill, 1990.
[20] H. M. Greenhouse, "Design of planar rectangular microelectronic inductors," *IEEE Trans. Parts, Hybrids, Packag.*, vol. 10, no. 2, pp. 101–109, Jun. 1974.
[21] R. G. Meyer, W. D. Mack, and J. J. E. M. Hageraats, "A 2.5 GHz BiCMOS transceiver for wireless LAN's," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 2097–2104, Dec 1997.
[22] B. Baytekin and R. G. Meyer. Analysis and simulation of spectral regrowth in linear RF power amplifiers. [Online]. Available: http://rfic.eecs.berkeley.edu/burcin/webpage/spectral.htm
[23] K. L. Fong and R. G. Meyer, "High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages," *IEEE J. Solid-State Circuits*, vol. 33, no. 4, pp. 548–555, Apr. 1998.
[24] K. L. Fong, "High-frequency analysis of linearity improvement technique of common-emitter transconductance stage using a low-frequency-trap network," *IEEE J. Solid-State Circuits*, vol. 35, no. 8, pp. 1249–1252, Aug. 2000.

Burcin Baytekin (S'97) was born in Ankara, Turkey, in 1975. He received the B.S. degree from the University of Southern California, Los Angeles, in 1997, and the M.S. and Ph.D. degrees from the University of California at Berkeley in 1999 and 2004, respectively, all in electrical engineering.

During the summer of 1998, he worked at Rockwell Semiconductor Systems, Newport Beach, CA, where he worked on a high-efficiency CDMA PA. During the summer of 1999, he worked at the IBM T. J. Watson Research Center, Yorktown Heights, NY, working on several PAs for Bluetooth transmitters. During the summers of 2000 and 2001, he worked at Maxim Integrated Products, Sunnyvale, CA, designing a high-dynamic-range peak detector for wireless communication applications and a 24-dBm 2.45-GHz Class-AB PA for IEEE802.11b, both of which went into production. Since 2003, he has been with Sequoia Communications, San Diego, CA, as an Analog/RF IC Designer.

Robert G. Meyer (S'64–M'68–SM'74–F'81) was born in Melbourne, Australia, on July 21, 1942. He received the B.E., M.Eng.Sci., and Ph.D. degrees in electrical engineering from the University of Melbourne in 1963, 1965, and 1968, respectively.

In 1968, he was employed as an Assistant Lecturer in electrical engineering at the University of Melbourne. Since September 1968, he has been employed in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, where he is now a Professor. His current research interests are high-frequency analog integrated-circuit design and device fabrication. He has acted as a consultant on electronic circuit design for numerous companies in the electronics industry. He is coauthor of the book *Analysis and Design of Analog Integrated Circuits* (Wiley, 1993) and Editor of the book *Integrated Circuit Operational Amplifiers* (IEEE Press, 1978).

Dr. Meyer was President of the IEEE Solid-State Circuits Council and was an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS and of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.

Simulation and Measurement of Supply and Substrate Noise in Mixed-Signal ICs

Brian E. Owens, Sirisha Adluri, Patrick Birrer, Robert Shreeve, *Member, IEEE*, Sasi Kumar Arunachalam, Kartikeya Mayaram, *Fellow, IEEE*, and Terri S. Fiez, *Fellow, IEEE*

***Abstract*—Digital noise in mixed-signal circuits is characterized using a scalable macromodel for substrate noise coupling. The noise coupling obtained through simulations is verified with measured data from a digital noise generator and noise sensitive analog circuits fabricated in the 0.35-μm heavily doped CMOS process. The simulations and measurements also demonstrate the effectiveness of including grounded guard rings and separating bulk and supply pins in digital circuits to reduce substrate coupling.**

***Index Terms*—Coupling noise, integrated circuit noise, mixed-signal noise, substrate coupling, substrate noise, supply noise.**

I. Introduction

THE integration of digital, analog, and RF circuitry to create systems-on-chips (SoCs) has become a reality in present day integrated circuits (ICs). Creating SoCs has major advantages including reduced size, reduced cost and lower power dissipation. However, this high level of complexity and integration causes noise coupling from the digital circuitry to the sensitive RF and analog circuitry. If the noise coupling is not addressed, it can result in significant performance degradation. The noise coupling occurs when the digital circuitry switches rapidly between high and low voltage levels. Current spikes are created that couple through the power supply and the shared silicon substrate [1]. Several approaches to modeling the substrate and simulating the digital noise have been developed [2]–[10]. Issues related to the proper inclusion of the package parasitics, backplane connections, and noise suppression techniques have not previously been adequately addressed.

This paper describes some of these issues and establishes guidelines for the simulation, measurement, and suppression of digital noise in mixed-signal integrated circuits. Section II presents background on the scalable macromodel used in this work [8], [11], [12]. This model serves as the foundation for validating the simulations with measurements. Section III separates out the contributions of substrate noise coupling due to supply noise and transistor switching. The package parasitics are shown to play a key role in the total substrate noise coupling in mixed-signal ICs. The digital noise generating circuitry and the analog sensing circuitry used to verify the circuit level noise simulations are described in Section IV. Section V presents measurement results from a test chip fabricated in a 0.35-μm heavily doped CMOS process and packaged in a 121-pin grid-array. The measurements validate that the simulation approach is very accurate. Based on the simulations and measurements, the effectiveness of various techniques for reducing the digital noise coupled into analog circuits is determined. Finally, Section VI concludes the paper.

Manuscript received January 6, 2004; revised August 14, 2004. This work was supported in part by SRC under Contract 2001-TJ-911 and by the DARPA TEAM project under Contract F33615-02-1-1179.

The authors are with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331 USA (e-mail: terri.fiez@eecs.oregonstate.edu).

Digital Object Identifier 10.1109/JSSC.2004.841039

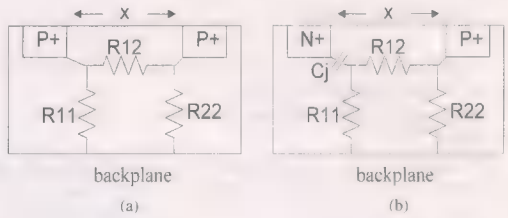


Fig. 1. Lumped substrate model for (a) p+ to p+ contacts and (b) n+ to p+ contacts.

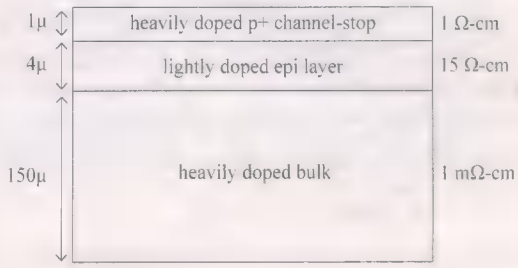


Fig. 2. Typical cross-section for a heavily doped substrate.

II. Substrate Coupling Macromodel

For efficient simulation of large SoCs, a simple model that accurately predicts substrate coupling must be used. Approaches including finite element methods [1], [15], [16], boundary element methods [3], [5], and polynomial curve fitting methods [17], [18] provide accurate post-layout simulation but they are computationally intensive particularly for full chip simulation. Additionally, they do not allow for pre-layout simulation. The substrate coupling model used in this work is scalable with contact shapes, dimensions, and separations [8], [11]. The substrate is modeled by a two-port lumped resistor network and it is valid for frequencies below a few gigahertz [2], [3]. The lumped resistive model for p+ to p+ contacts and n+ to p+ contacts is shown in Fig. 1(a) and (b), respectively. The resistance, R12, models the coupling between the two contacts and R11 and R22 model the coupling from the contacts to the backplane. The n+

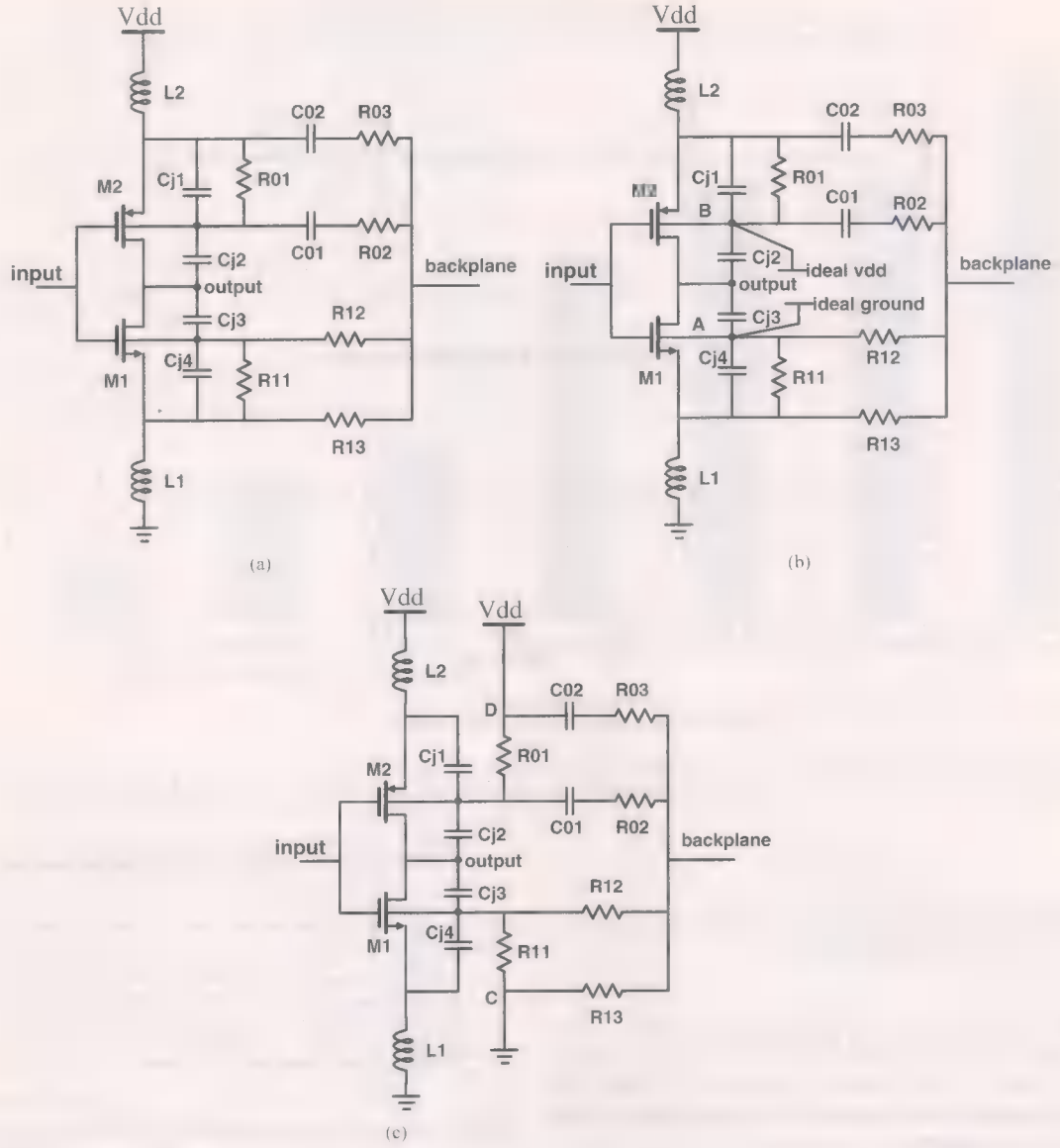


Fig. 3. (a) Setup to measure the total noise at the substrate. (b) Setup used to measure supply-only noise at the substrate. (c) Setup used to measure switching-only noise at the substrate.

contact to p-type substrate junction capacitance is modeled by Cj.

The resistance values are determined by characterizing the substrate either through device simulations such as with the Medici simulator [19] or through measurements of the substrate. A typical heavily doped substrate profile is shown in Fig. 2 and consists of three distinct layers: a heavily doped p+ channel-stop implant, a lightly doped epitaxial (epi) layer, and a heavily doped p+ bulk [3]. The layer resistivities and thicknesses determine the substrate coupling (and resistance values) between contacts and to the backside.

As the separation between contacts increases in heavily doped processes, the resistance between the contacts becomes very large. At separations beyond about 100 μm, nearly all of the current from the digital noise sources flows down into the substrate through the resistance to the backplane and then back up into the analog circuits when the backplane is floating. For this reason, if the separation between digital and analog circuitry is greater than 100 μm, increasing the separation beyond this point provides only negligible improvement in the substrate coupling. Table I illustrates typical resistor values for two identical contacts at various separations in a heavily doped substrate. Notice that if these resistor values are used in the model shown in Fig. 1 and the backplane is left floating, the resistors R_{11} and R_{22} are indeed the dominant contact coupling path.

TABLE I
SAMPLE SUBSTRATE RESISTANCES IN A HEAVILY DOPED CMOS PROCESS

Separation (μm)	R_{11} (Ω)	R_{22} (Ω)	R_{12} (Ω)
10	390	390	962
50	305	305	33.8K
100	303	303	2.58M

The scalable macromodel is based on Z-parameters from which the resistances can be derived or *vice versa* [12]. The

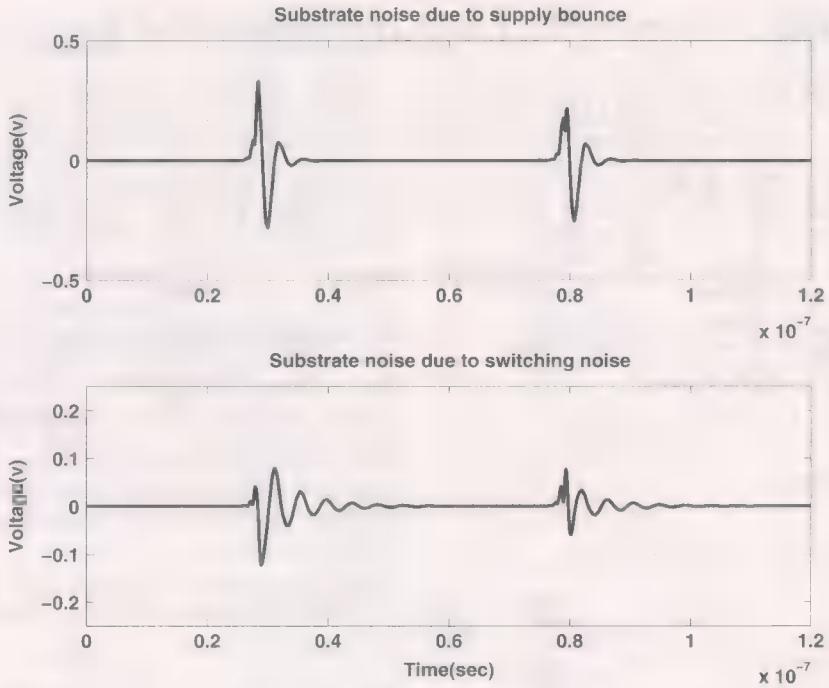


Fig. 4. Simulated plots of supply noise (top) and switching noise (bottom) for a stepped buffer.

model is expressed in terms of Z_{11}, Z_{12}, and Z_{22}. Z_{12} is given by

$$Z_{12} = \alpha e^{-\beta x}$$

where x is the distance between contacts and α and β are process parameters. Z_{11} (Z_{22}) is given by

$$Z_{11} = \frac{1}{K_1 \text{Area} + K_2 \text{Perimeter} + K_3}$$

where K_1, K_2, and K_3 are process parameters. Using this macromodel, substrate resistances can be obtained for an arbitrary number of contacts.

III. Dependence of Power Supply and Transistor Switching Noise on Package Parasitics

The power supply noise and transistor switching noise of a seven stage stepped buffer are simulated to illustrate the contributions of each to the substrate coupling noise. Stepped buffers are often a major source of substrate and supply noise in mixed-signal ICs as they provide buffering for clock signals as well as for output buffers to drive large off-chip capacitance. The stepped buffer consists of seven stages of inverters, with each successive stage loaded by two inverters sized a power of e larger than the previous, one is part of the seven stage stepped buffer and the other serves as a dummy load. The first inverter transistors are sized $(W/L)\text{n} = 5\ \mu\text{m}/0.6\ \mu\text{m}$ and $(W/L)\text{p} = 10\ \mu\text{m}/0.6\ \mu\text{m}$. The stepped buffer is designed and laid out in a 0.35-μm heavily doped CMOS process and the resistive substrate network is extracted for this process and design. An inductance of 5 nH is included in the power and ground lines to model the effect of the bond wire and package inductance. The supply and switching noise generated by the stepped buffer are simulated using the approach in [7] and using the circuits shown in Fig. 3. Fig. 3(a) is the equivalent circuit used to simulate both the supply and substrate noise contribution. The substrate resistances are extracted from the macromodel or a boundary element solver. Figs. 3(b) and (c) are the equivalent circuits to simulate the supply-only and substrate-only coupling. The substrate voltage is measured at the backplane node so that there is no dependence on the substrate contact size. The simulated results are shown in Fig. 4. The peak-to-peak value of the supply noise is three times larger than the peak-to-peak value of the switching noise. The dominance of the supply noise over the switching noise is due to the presence of the large supply inductance.

TABLE II
Range of Pin Parasitics for Different Packages [20], [21]

Type	No of pins	Self Inductance (nH)	Capacitance (pF)	Resistance (mΩ)
DIP	40	4 to 9	0.6 to 5	24 to 200
PGA	108	3 to 7	1.5 to 3	100 to 400
SOIC	44	4 to 12	0.1 to 0.2	200 to 1000
BGA	225	7.5 to 14.5	0.1 to 0.2	200 to 450
LPCC	32	1.1 to 1.47	.08 to 0.11	165 to 190
Flip chip	64	0.26 to 1.5	0.18 to 0.38	7 to 54

The type of package used in the design of mixed-signal ICs and its particular parasitics can have a profound effect on the substrate and supply noise coupling. Several packages and their associated parasitic capacitances, resistances and inductances are illustrated in Table II. The stepped buffer is again simulated using the average values for each package in Table II and the results are shown in Fig. 5. When the rms value of the substrate noise is compared in all cases, it can be seen that the rms noise varies from 3 mV for no package model to 24 mV for the BGA which is a factor of 8 times larger. When the stepped buffer is simulated with the flip-chip and LPCC package models, the substrate noise is a factor of 3 lower than the BGA case. On-chip

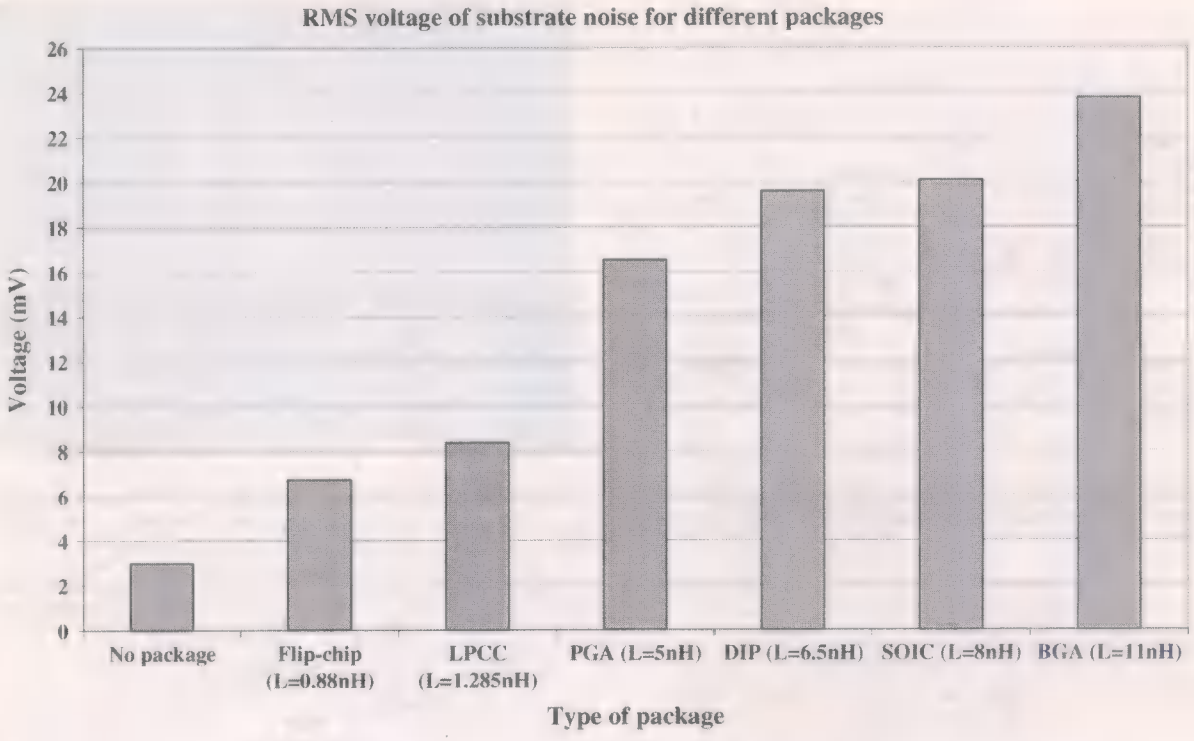


Fig. 5. Graph showing the rms values of substrate noise generated by the stepped buffer for different package parasitics. The input frequency of the stepped buffer is 780 kHz. Source and bulk nodes of the transistors in the stepped buffer are connected to separate supplies.

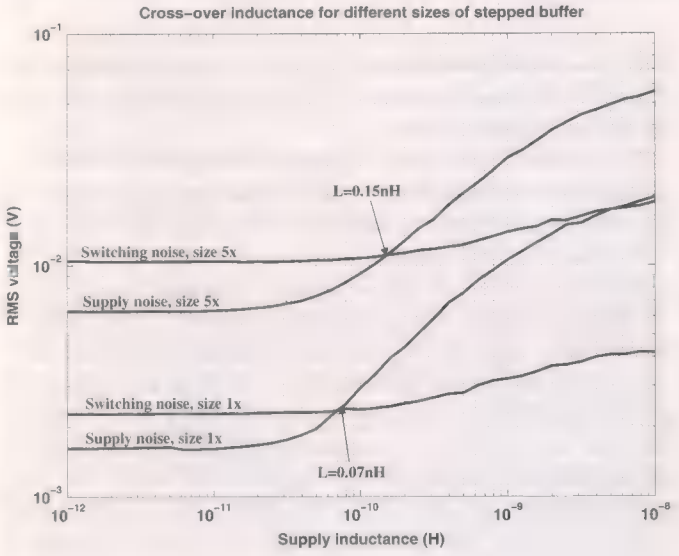


Fig. 6. Affect of transistor sizing on the substrate noise coupling as a function of the supply inductance. The input frequency of the buffer is 10 MHz. $(W/L)_{n1x} = 1\ \mu$m/0.6 μm, $(W/L)_{p1x} = 2\ \mu$m/0.6 μm, $(W/L)_{n5x} = 5\ \mu$m/0.6 μm, $(W/L)_{p5x} = 10\ \mu$m/0.6 μm.

decoupling capacitance can also be used to significantly reduce the supply noise.

The value for which the supply noise dominates over the transistor switching noise in a stepped buffer changes as the size of the stepped buffer changes. A second seven stage stepped buffer is designed that is one-fifth the size of the previous buffer. Both are simulated as a function of the supply inductance with an input frequency of 10 MHz and the results are plotted in Fig. 6.

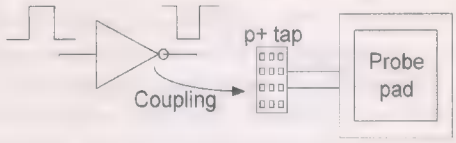


Fig. 7. Measurement setup used for directly probing the substrate. Noise was measured via a p+ substrate tap connected directly to a probe pad.

For the larger stepped buffer, the transistor switching noise dominates up to a supply inductance of 0.15 nH compared to 0.07 nH for the smaller stepped buffer. However, it is also important to note the substrate coupling noise is nearly an order of magnitude higher for the large stepped buffer. This indicates that for the packages commonly used, supply noise will be the dominant contributor to the noise.

IV. Measurement of Substrate Noise

Three different noise-sensing methods were used to characterize the substrate coupling. The first method uses p+ substrate taps connected to pads that can be directly probed as shown in Fig. 7. This method is the simplest because it does not require additional on-chip circuitry for the measurement; however, it is generally not as accurate since the probe impedance may load the substrate.

In the second measurement approach, a wide-band differential output amplifier based on the design in [7] is used to buffer the substrate from the probe impedance. The amplifier, shown in Fig. 8, has one input connected to the substrate via a large MOS capacitor and the other input is connected to a separate quiet bias voltage. The input MOS capacitor is quite large, so at the frequencies of interest it acts as a short circuit. The amplifier has

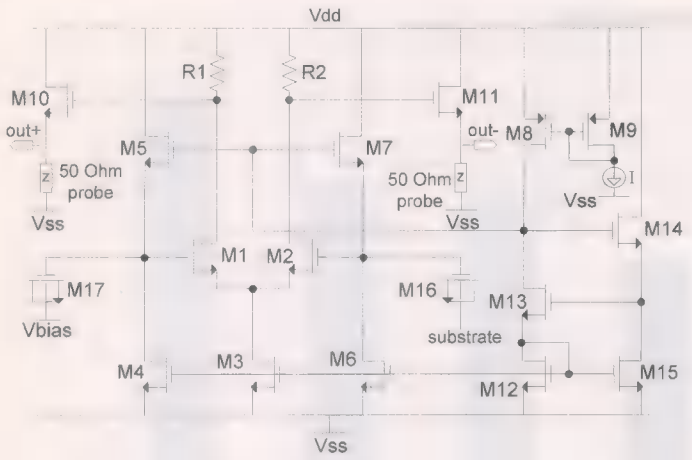


Fig. 8. Noise-sensing amplifier used to measure the noise in the substrate. One input is sensitive to substrate noise while the other is connected to a quiet ground.

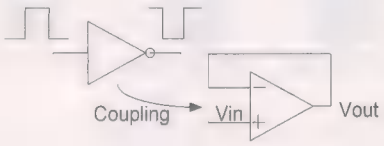


Fig. 9. Noise coupling measurement setup for the folded-cascode amplifier connected in a unity-gain configuration.

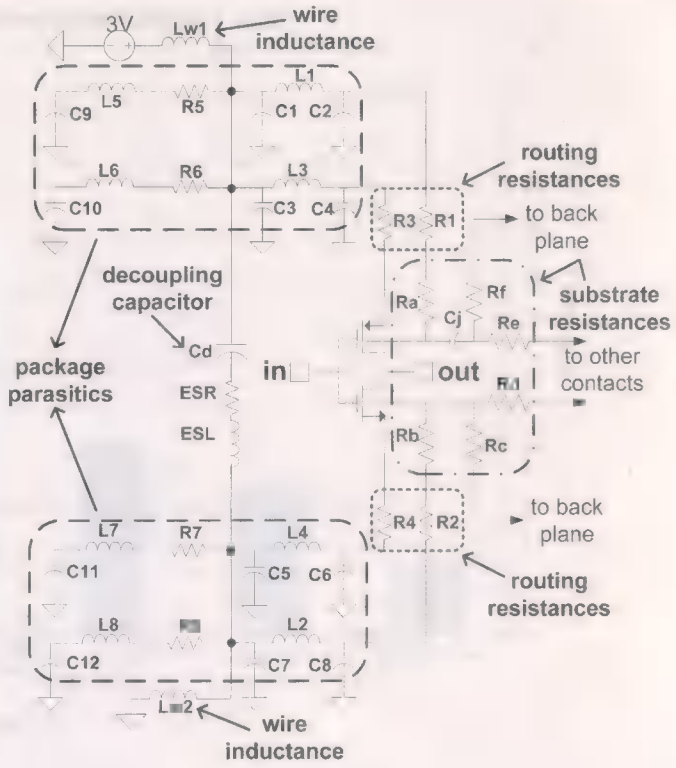


Fig. 10. Diagram depicting one stage of the digital stepped buffer showing the parasitics and the substrate network. Separate source and bulk power supplies were used.

been designed so that a 50-Ω impedance high-frequency probe can be connected at its output without changing the overall performance. Additionally, the amplifier is designed for a 700-MHz bandwidth, making it possible to perform continuous-time measurements of the substrate at high frequencies. The probes used to measure the output of the buffer amplifier were high bandwidth ground–signal–ground probes. Although the gain of the buffer amplifier is relatively low, approximately 3 dB, this does not limit the overall measurement as long as the amplitude of the digital noise is within the range of the measurement accuracy.

The noise-sensing buffer amplifier layout was arranged to maximize the matching between the input transistors and load resistors so that the input-referred offset is minimized and the maximum amount of substrate noise is sensed. This is achieved by interdigitating the input transistors and load resistors and incorporating dummy transistors and capacitors into the arrays. Additionally, the common-mode rejection ratio (CMRR) is maximized in the design. Ground–signal–ground pads were placed above and below the opamp itself. This enables the routing traces from the circuit to the pads to be as short as possible, but still spaced far enough apart to meet the probing requirements.

The third and final substrate noise sensing method involved the use of an analog building block that in this case was a folded-cascode amplifier connected in a unity-gain configuration [14]. The purpose of the operational amplifier is to demonstrate the application of the model and simulation approach for evaluating simple mixed-signal circuits. A block diagram illustrating the setup is shown in Fig. 9. By clocking the digital circuitry, noise is injected into the substrate and the power supply lines. This noise couples into the bulks of the input transistors and the supply lines if they are shared. A seven stage stepped buffer, described in the previous section, is the noise injecting source for all the measurement cases.

To accurately predict the digital noise coupled into the substrate, it is imperative that parasitic capacitances, inductances, and resistances associated with the package and the connection to the backplane be included in the overall simulation. Fig. 10 shows the inclusion of critical parasitic elements as well as the substrate resistances for one inverter stage of a digital stepped buffer. These parasitic elements match those used for measurement of the test chip as described later. The package parasitics are obtained from the package model for a 121-pin grid-array package, whereas the substrate resistances were extracted using the scalable macromodel and are indicated by Ra-Rf. On-chip interconnect resistances, R1–R4, are also modeled and included in the simulations. An off-chip decoupling capacitor, Cd (and its parasitics ESR and ESL), is used in the actual measurements to reduce the supply bounce.

V. Experimental Results

A test chip was fabricated in a 0.35-μm heavily doped CMOS quad-metal, double poly process. It consisted of a stepped buffer, a folded-cascode amplifier connected in unity gain and two substrate noise sensing amplifiers as shown in Fig. 11. A single 3-V supply was used for all measurements. The stepped buffer was placed approximately 100 μm away from the folded-cascode amplifier and 400 μm and 800 μm away from the two noise-sensing amplifiers. At distances above approximately 100 μm, the cross-coupling resistance between

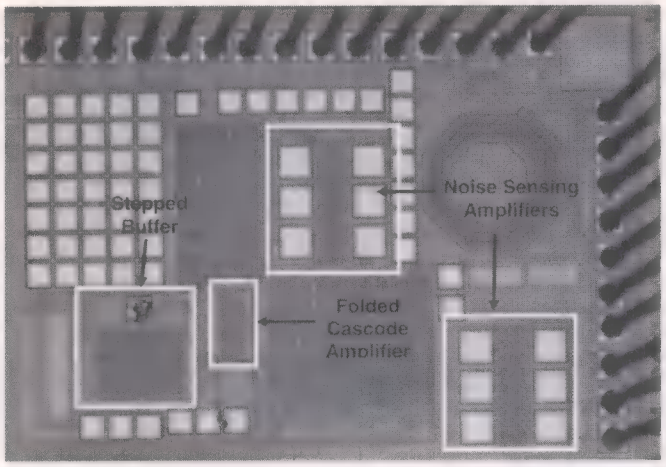


Fig. 11. Microphotograph of the test chip containing the folded-cascode amplifier, noise-sensing amplifiers, and stepped buffer.

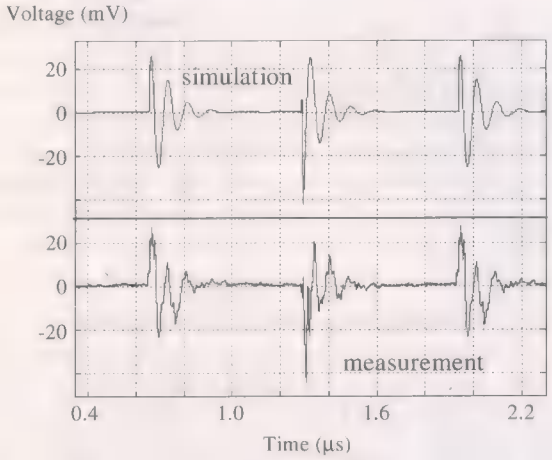


Fig. 12. Simulation and measurement of the substrate noise picked up by the noise-sensing amplifier with the stepped buffer running at a frequency of 780-kHz.

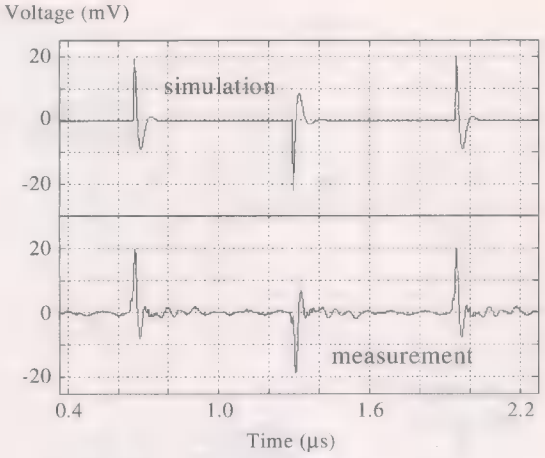


Fig. 13. Simulation and measurement of the substrate noise sensed at the substrate tap with the stepped buffer running at a frequency of 780 kHz.

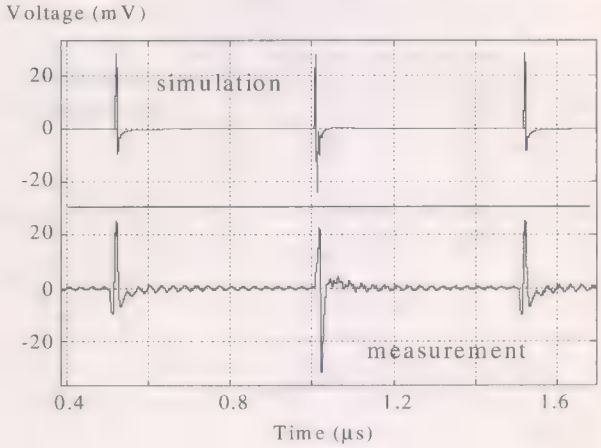


Fig. 14. Simulation and measurement of the substrate noise sensed by the folded-cascode amplifier connected in a unity-gain configuration. The stepped buffer was operating at 1 MHz.

circuit elements is so large (in the MΩ range), that almost all the noise is transmitted down into the substrate and then back up into the circuit elements when the backplane is floating. For this reason, all the measurements from both of the noise-sensing amplifiers were identical.

The transient behavior was measured using all three measurement techniques previously described. The noise-sensing amplifier provided a means by which continuous time measurements of the substrate could be made without loading the measurement with the probe impedance.

Shown in Fig. 12 are the simulations and measurements at the output of the noise-sensing amplifier when the stepped buffer is clocked with a 3.3-V 780-kHz input waveform. The relative voltage peaks and the amount of ringing from both the simulations and the measurements are in good agreement. Fig. 13 shows simulations and measurements made using the p+ substrate tap as the means of measuring noise from the stepped buffer. In contrast to the results from the buffer amplifier, the substrate tap measurement amplitude is smaller and has less ringing due to the loading of the probe. In Fig. 14 the measurement and simulation of the folded-cascode amplifier output in the unity-gain configuration is shown when the stepped buffer is clocked. Once again, the general shape and peak-to-peak voltage amplitude match very closely.

A. Separating Supply and Bulk Connections in Digital Circuits

When supply noise is dominant in digital circuits, it may be possible to reduce the noise by separating the transistor's bulk and source power supplies. Under normal circumstances, a transistor in a digital circuit will have its bulk and source nodes tied together on chip and taken to a single pin. By using separate pins for the bulks and sources, voltage bounce on the power supply lines is not fed directly into the transistor bulks. This may help reduce the amount of noise which is injected into the substrate. Figs. 15(a) and (b) illustrate the two different scenarios, where the bulks and sources are connected to a single pin, Fig. 15(a), and to separate pins, Fig. 15(b).

Fig. 16 shows the simulation and measurement results for the noise picked up by the substrate tap when the stepped buffer's bulks and sources are connected as shown in Fig. 15(a). By tying the sources and bulks together on chip, the peak-to-peak noise picked up approximately doubles as shown in Fig. 16(b). Our results are different from those in [22] due to the smaller amount

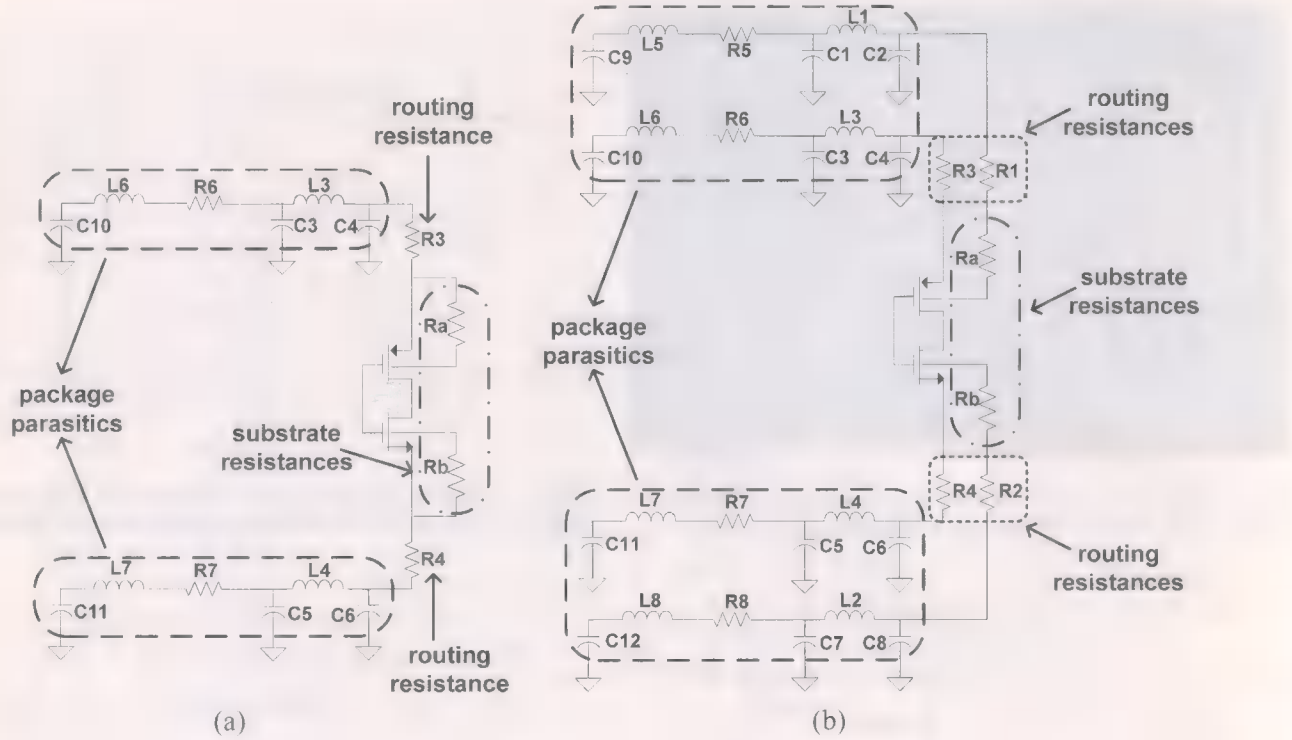


Fig. 15. Digital bulk and source power supply routing using (a) one pin and (b) using two pins.

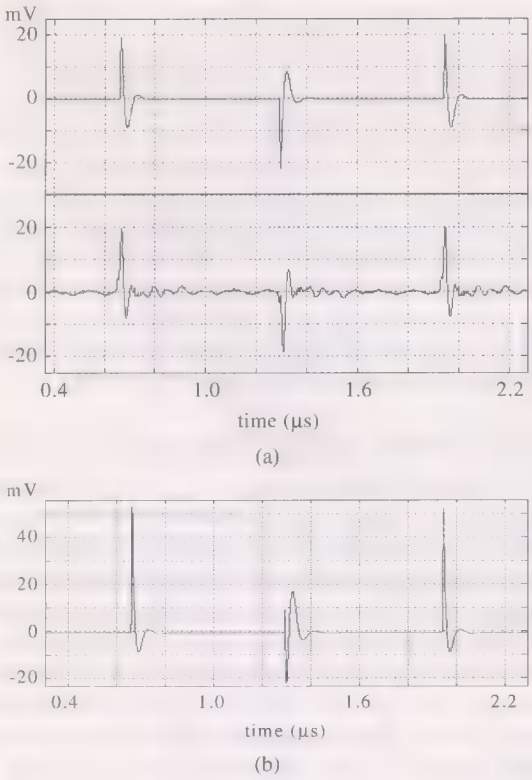


Fig. 16. (a) Simulation (top) and measurement (bottom) of the noise picked up by the substrate tap when the stepped buffer's bulks and sources are connected to separate pins. (b) Simulation of the noise at the substrate tap when the bulks and sources are tied together and routed to a single pin.

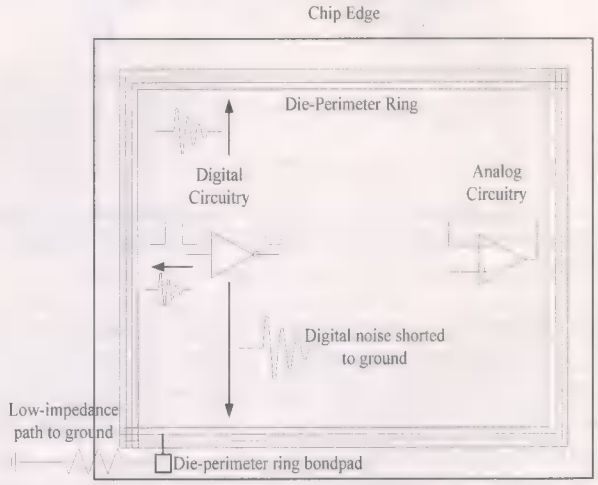


Fig. 17. Conceptual representation of the die-perimeter ring connections to reduce the noise coupling through the substrate.

of digital circuitry in the test chip. For chips dominated by digital circuitry, separating the transistor sources and bulks may not necessarily improve the substrate coupling noise as shown in [22].

At higher frequencies, the source-to-bulk capacitances can create an effective short. In the test circuit shown in this paper, the capacitance for the largest transistor is 0.8 pF which translates into a resistance of 2 kΩ at a frequency of 100 MHz. Since, the routing resistance is only about 50 Ω, this capacitor will not provide an on-chip short between the source and the bulk even for frequencies up to several hundred megahertz.

Interestingly, the negative noise peaks remain relatively unaffected by the change, while the positive spikes triple in amplitude for the case where the source and bulk nodes are tied

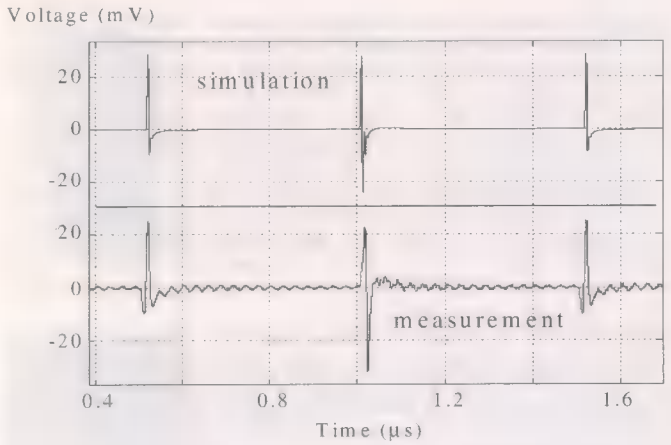


Fig. 18. Simulation (top) and measurement (bottom) of the folded-cascode amplifier with the die-perimeter ring left floating.

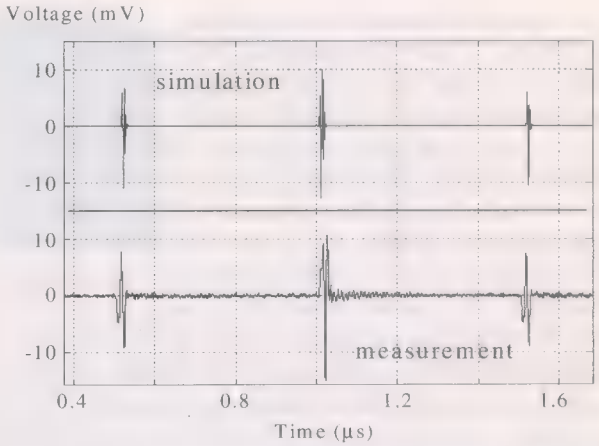


Fig. 19. Simulation (top) and measurement (bottom) of the folded-cascode amplifier with the die-perimeter ring grounded.

together. The reason for this behavior was a large off-chip capacitance that is being driven by the stepped buffer. Because the noise injected in the substrate is determined by the discharge of this large capacitance, this does not change when the bulk and supplies are separated. For smaller values of off-chip capacitance (or no capacitance) both peaks are reduced by separating the bulks and the power supplies.

B. Grounding the Die Perimeter Ring

When switching noise is dominant in heavily doped substrates, it can be reduced by grounding the substrate backside. Backside metallization is one method of grounding the backplane. However, this is not standard and adds extra cost to production. On the other hand, die-perimeter rings are standard on many chip designs since they are often used in electrostatic discharge (ESD) protection schemes to reduce the ground resistance between I/O pads. On this test chip, a grounded die-perimeter ring has been used to ground the backplane since the measured resistance between the backplane and the die-perimeter ring is only 1.6 Ω. A schematic of the setup is shown in Fig. 17.

Figs. 18 and 19 show the simulations and measurements of the folded-cascode amplifier with and without the die-perimeter ring grounded [13]. The die-perimeter ring resistance to the backplane was measured to be 1.6 Ω and this was used in the simulation. In both the simulations and measurements without die-perimeter ring grounding, the peak-to-peak noise voltage observed is around 55 mV. Contrasting this to the noise voltage of the grounded case, it can be seen that the peak-to-peak value is approximately halved and is now at 26 mV.

Simulations shown in Fig. 20 summarize the effects of grounding the backplane and separating the bulk and supply connections for the PMOS devices. Separating the bulk and source nodes reduces the substrate noise by approximately 7.5 dB at high inductance values, e.g., 1 nH. When the bulk and source nodes are tied together and the backplane is grounded, the substrate noise reduction on the backplane is 8.5 dB for low inductance values. There is approximately a 15-dB reduction in substrate noise when the source and bulk nodes of the transistors are tied separately and when the backplane is grounded,

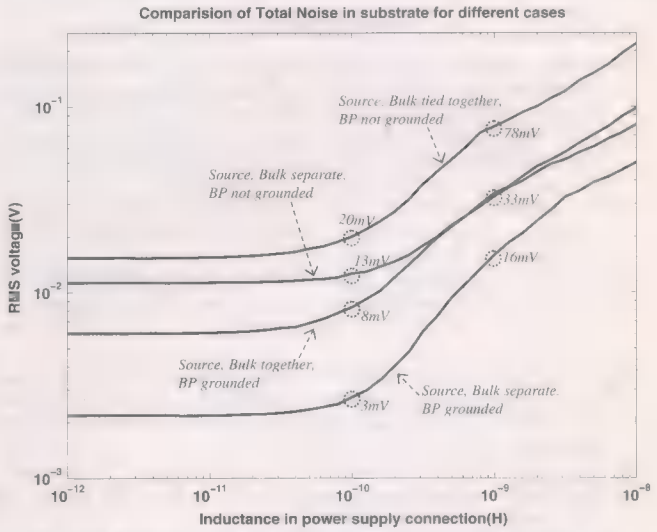


Fig. 20. Comparison of total substrate noise generated by the stepped buffer for different cases: bulk and source nodes tied together and separate, back plane grounded and not grounded. The input frequency of the buffer is 10 MHz.

compared to connecting the source and bulk nodes and floating the backplane.

Grounding the die-perimeter ring is an effective way to reduce digital noise. However, the die-perimeter ring must have a low-impedance path to ground to be effective. This can be achieved by connecting the die-perimeter ring to a bond-pad and then down-bonding. If a high impedance path (i.e., a long wire with significant inductance) is used to connect the die-perimeter ring to ground, the results may show no change or even an increase in noise coupling.

VI. Conclusion

An approach for simulating digital noise coupling has been discussed and verified using measurements from a test-chip fabricated in a 0.35-μm heavily doped CMOS process. Measurements were shown for noise coupled from a stepped buffer to an analog noise-sensing amplifier, folded-cascode amplifier, and substrate tap. These measurements match closely with the simulated results. Based on these measurements and simulations it

can be concluded that the macromodel gives a good approximation of the noise that will be coupled to a given analog circuit. Additionally, noise suppression techniques have also been discussed. Measurements and simulations for our test chip show that more than 6 dB noise reduction can be achieved by using separate digital bulk and source power supply pins, and more than 6-dB reduction can be obtained by using a die-perimeter ring connected to ground.

The results of this work show that the choice of packages for mixed-signal chips greatly affects the amount of total substrate noise. Below package inductance values of 100 pH, noise coupling from MOSFETs in the cases presented here is dominant. Further reduction in package inductance beyond this point will only slightly reduce the total substrate noise generated. Most flip chip packages satisfy this criteria and are therefore an excellent choice for ensuring power supply noise coupling is not a factor.

REFERENCES

[1] N. K. Verghese, T. J. Schmerbeck, and D. J. Allstot, *Simulation Techniques and Solutions for Mixed-Signal Coupling in Integrated Circuits*. Boston, MA: Kluwer, 1995.

[2] N. K. Verghese and D. J. Allstot, "Computer-aided design considerations for mixed-signal coupling in RF integrated circuits," *IEEE J. Solid-State Circuits*, vol. 33, no. 3, pp. 314–323, Mar. 1998.

[3] R. Gharpurey and R. G. Meyer, "Modeling and analysis of substrate coupling in integrated circuits," *IEEE J. Solid-State Circuits*, vol. 31, no. 3, pp. 344–353, Mar. 1996.

[4] D. K. Su, M. J. Loinaz, S. Masui, and B. A. Wooley, "Experimental results and modeling techniques for substrate noise in mixed-signal integrated circuits," *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 420–430, Apr. 1993.

[5] J. P. Costa, M. Chou, and L. M. Silveira, "Efficient techniques for accurate modeling and simulation of substrate coupling in mixed-signal IC's," *IEEE Trans. Computer Aided Design*, vol. 18, no. 5, pp. 597–607, May 1999.

[6] A. J. van Genderen, N. P. van der Meijs, and T. Smedes, "Fast computation of substrate resistances in large circuits," in *Proc. Eur. Design and Test Conf.*, Mar. 1996, pp. 560–565.

[7] M. van Heijningen, J. Compiet, P. Wambacq, S. Donnay, M. G. E. Engels, and I. Bolsens, "Analysis and experimental verification of digital substrate noise generation for epi-type substrates," *IEEE J. Solid-State Circuits*, vol. 35, no. 7, pp. 1002–1008, Jul. 2000.

[8] A. Samavedam, A. Sadate, K. Mayaram, and T. S. Fiez, "A scalable substrate noise coupling model for design of mixed-signal IC's," *IEEE J. Solid-State Circuits*, vol. 35, no. 6, pp. 895–904, Jun. 2000.

[9] M. Badaroglu, M. van Heijningen, V. Gravot, J. Compiet, S. Donnay, G. G. E. Giellen, and H. J. De Man, "Modeling and experimental verification of substrate noise reduction in CMOS mixed-signal IC's with synchronous digital circuits," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1383–1395, Nov. 2002.

[10] M. Badaroglu, S. Donnay, H. J. DeMan, Y. A. Zinzius, G. G. E. Giellen, W. Sansen, T. Fonden, and S. Signell, "Modeling and experimental verification of substrate noise generation in a 220-kgates WLAN system-on-chip with multiple supplies," *IEEE J. Solid-State Circuits*, vol. 38, no. 7, pp. 1250–1260, Jul. 2003.

[11] H. D. Ozis, T. Fiez, and K. Mayaram, "A comprehensive geometry-dependent macromodel for substrate noise coupling in heavily doped CMOS processes," in *Proc. IEEE Custom Integrated Circuits Conf.*, May 2002, pp. 497–500.

[12] H. D. Ozis, "An efficient modeling approach for substrate noise coupling analysis with multiple contacts in heavily doped CMOS processes," M.S. Thesis, Oregon State Univ., Corvallis, OR, Aug. 2001.

[13] B. Owens, P. Birrer, S. Adluri, R. Shreeve, S. K. Arunachalam, H. Habal, S. Hsu, A. Sharma, K. Mayaram, and T. S. Fiez, "Strategies for simulation, measurement, and suppression of digital noise in mixed-signal circuits," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2003, pp. 361–364.

[14] A. C. Sadate, "A substrate noise coupling model for lightly doped CMOS processes," M.S. Thesis, Oregon State Univ., Corvallis, OR, Dec. 2000.

[15] B. R. Stanisic, N. K. Verghese, R. A. Rutenbar, R. L. Carley, and D. J. Allstot, "Addressing substrate coupling in mixed-mode IC's and power distribution synthesis," *IEEE J. Solid-State Circuits*, vol. 29, no. 3, pp. 226–237, Mar. 1994.

[16] I. L. Wemple and A. T. Yang, "Integrated circuit substrate coupling models based on Voronoi tessellation," *IEEE Trans. Computer-Aided Design*, vol. 14, no. 12, pp. 1459–1468, Dec. 1995.

[17] N. K. Verghese, D. J. Allstot, and M. A. Wolfe, "Fast parasitic extraction for substrate coupling in mixed-signal ICs," in *Proc. IEEE Custom Integrated Circuits Conf.*, Feb. 1995, pp. 121–124.

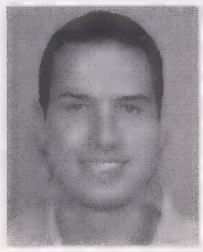
[18] ——, "Verification techniques for substrate coupling and their applications to mixed-signal IC design," *IEEE J. Solid-State Circuits*, vol. 31, no. 3, pp. 354–365, Mar. 1996.

[19] *MEDICI, Version 200.2.0*, 2000. Avant! Corp..

[20] (2003, Aug.) IC Package Electrical Characteristics. Mitsubishi Integrated Circuit Packages. [Online]. Available: http://www.mitsubishi.com/

[21] (2003, Aug.) Ceramic Packaging Options. The Mosis Service. [Online]. Available: http://www.mosis.org/Technical/Packaging/Ceramic/menu-pkg-ceramic.html

[22] M. Felder and J. Ganger, "Analysis of ground-bounce induced substrate noise coupling in a low resistive bulk epitaxial process: Design strategies to minimize noise effects on a mixed-signal chip," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 11, pp. 1427–1436, Nov. 1999.



Brian E. Owens received the B.S. (*magna cum laude*) and M.S. degrees in electrical engineering from Oregon State University, Corvallis, in 2001 and 2003, respectively.

Since October 2003, he has been with Sandia National Laboratories, Livermore, CA, where he is a Member of the Technical Staff in the Reliability and Electrical Systems Department.

Sirisha Adluri received the B.E (Hons.) degree from the Birla Institute of Technology and Science, Pilani, India, in 2000 and the M.S degree from Oregon State University, Corvallis, in 2003, both in electrical engineering.

She is currently with Texas Instruments Inc. as a Mixed-Signal Circuit Design Engineer. Her interests include analog and mixed-signal integrated circuit design.

Patrick Birrer received the B.S. degree in electrical engineering from the Burgdorf School of Engineering, Switzerland, in 2000 and the M.S. degree in electrical engineering from Oregon State University, Corvallis, in 2004.

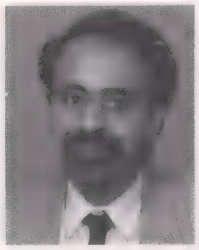
He joined Cadence Design Systems Germany in 2004. His interests include analog/mixed signal IC design and simulation, behavioral modeling, and formal verification.

Robert Shreeve (S'79–M'81) received B.S. degrees in electrical engineering and chemistry from the University of Idaho, Moscow, in 1981. His is currently working toward the Ph.D. degree at Oregon State University, Corvallis.

His research interest is analytical modeling of substrate coupling mechanisms. He has worked for Hewlett Packard since 1985 specializing in IC process development, IC circuit design, and MEMs research.

Sasi Kumar Arunachalam received the B.E.(Hons.) degree from the Birla Institute of Technology and Science, Pilani, India, in 2002. He is currently working toward the M.S degree at Oregon State University, Corvallis.

He held an internship position at Qualcomm Incorporated, San Diego, CA, in summer 2003, where he analyzed substrate coupling in mixed-signal designs. His research interests are analog and mixed-signal integrated circuit design.



Kartikeya Mayaram (S'82–M'88–SM'99–F'05) received the B.E. (Hons.) degree in electrical engineering from the Birla Institute of Technology and Science, Pilani, India, in 1981, the M.S. degree in electrical engineering from the State University of New York, Stony Brook, in 1982, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 1988.

From 1988 to 1992, he was a Member of Technical Staff in the Semiconductor Process and Design Center of Texas Instruments, Dallas. From 1992 to 1996, he was a Member of Technical Staff at Bell Laboratories, Allentown, PA. He was an Associate Professor in the School of Electrical Engineering and Computer Science, Washington State University, Pullman, from 1996 to 1999 and in the Electrical and Computer Engineering Department at Oregon State University, Corvallis, from 2000 to 2003. Now, he is a Professor in the School of Electrical Engineering and Computer Science at Oregon State University. His research interests are in the areas of circuit simulation, device simulation and modeling, integrated simulation environments for microsystems, and analog/RF design.

Dr. Mayaram received the National Science Foundation (NSF) CAREER Award in 1997. He was an Associate Editor of IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS from 1995 to 2001 and has been the Editor-in-Chief of this journal since January 2002.

Terri S. Fiez (S'82–M'85–SM'95–F'05) received the B.S. and M.S. degrees in electrical engineering from the University of Idaho, Moscow, in 1984 and 1985, respectively. In 1990, she received the Ph.D. degree in electrical and computer engineering from Oregon State University, Corvallis.

From 1985 to 1987 and in 1988 she worked at Hewlett-Packard Corporation in Boise, ID, and Corvallis, OR, respectively. In 1990, she joined Washington State University, Pullman, as an Assistant Professor, where she became an Associate Professor in 1996. In the fall of 1999, she joined the Department of Electrical and Computer Engineering at Oregon State University as a Professor and Department Head. She became Director of the School of Electrical Engineering and Computer Science in 2003. Her research interests are in the design of high-performance analog signal processing building blocks, simulation and modeling of substrate coupling effects in mixed-signal ICs, and innovative engineering education approaches.

Dr. Fiez has been involved in a variety of IEEE activities, including serving on the committees for the IEEE International Solid-State Circuits Conference, IEEE Custom Integrated Circuits Conference, ISCAS, and as a guest editor of the JOURNAL OF SOLID-STATE CIRCUITS. She was previously awarded the NSF Young Investigator Award and the Solid-State Circuit Predoctoral Fellowship.

High-Performance RF Mixer and Operational Amplifier BiCMOS Circuits Using Parasitic Vertical Bipolar Transistor in CMOS Technology

Ilku Nam, *Student Member, IEEE,* and Kwyro Lee, *Senior Member, IEEE*

***Abstract*—The electrical characteristics of the parasitic vertical NPN (V-NPN) BJT available in deep n-well 0.18-μm CMOS technology are presented. It has about 20 of current gain, 7 V of collector-emitter breakdown voltage, 20 V of collector-base breakdown voltage, 40 V of Early voltage, about 2 GHz of cutoff frequency, and about 4 GHz of maximum oscillation frequency at room temperature. The corner frequency of $1/f$ noise is lower than 4 kHz at 0.5 mA of collector current. The double-balanced RF mixer using V-NPN shows almost free $1/f$ noise as well as an order of magnitude smaller dc offset compared with CMOS circuit and 12 dB flat gain almost up to the cutoff frequency. The V-NPN operational amplifier for baseband analog circuits has higher voltage gain and better input noise and input offset performance than the CMOS ones at the identical current. These circuits using V-NPN provide the possibility of high-performance direct conversion receiver implementation in CMOS technology.**

***Index Terms*—BiCMOS, deep n-well CMOS, direct conversion receiver, offset, operational amplifier, parasitic vertical bipolar transistor, RF mixer, $1/f$ noise.**

I. Introduction

COMPARED with MOSFET, the BJT (Bipolar Junction Transistor) devices have many desirable characteristics for analog applications including RF, namely, much smaller $1/f$ noise, much better device-to-device matching, larger transconductance, easier biasing, and easier impedance matching, and so forth. For this reason, RF and analog circuit designers usually prefer the use of BJT over MOSFET and most state-of-the-art radio chips have been fabricated using BiCMOS processes where the high performance vertical Si/Ge BJT is used for RF circuit and CMOS for logic [1]–[3]. However, the BiCMOS process has several drawbacks that the cost is expensive, the period of process development is long, the foundry service is very limited, and the performance of BiCMOS digital circuits is inferior to that of CMOS ones. As a result, this process may be unsuitable for the implementation of low cost single chip radio.

On the other hand, continuous advances in CMOS technology provide both good RF circuits and digital VLSI at very low cost [4], [5]. Deep submicron CMOS process has been regarded very plausible to integrate digital modem blocks. In modern wireless communication receivers, highest degrees of integration are achieved with the direct conversion receiver (DCR). Therefore, the DCR's realization in CMOS technology has extensively been studied as a possible solution for low cost single-chip radio [6], [7]. However, CMOS DCR has the inherently serious problems of $1/f$ noise, dc offset, *I/Q* mismatch, LO (local oscillator) leakage, even order distortion, and so on [8]. Although, some of these can be alleviated by using novel circuit technique, careful layout, and compensation by digital signal processing, the $1/f$ noise and dc offset problems have been critical issues in CMOS analog circuits because MOSFET device has very large $1/f$ noise and mismatch in itself. These are especially problematic for DCR and baseband analog (BBA) circuits, which seriously degrade the overall sensitivity of CMOS receiver and raise an obstacle to its commercialization.

Therefore, there have been many trials to use parasitic lateral BJT available in CMOS technology [9]–[14]. Because its base width is basically determined by the MOSFET gate length, very high current gain and unit current gain cutoff frequency are expected from scaled down CMOS technology. However, the uniformity, reproducibility, device matching, and driving capability of these lateral devices are very questionable to be useful for practical purpose. In addition, there has been some effort to make use of the parasitic substrate vertical BJT available in double-well CMOS process [15]. However, the use of this transistor is very limited since its collector is tied together to the substrate. Moreover, its RF performance is not satisfactory because of thick well depth.

In this paper, we present the RF characteristics of parasitic vertical NPN (V-NPN) BJT available in deep n-well CMOS process [16] and the result of utilizing the V-NPN for low $1/f$ noise and dc offset RF mixer as well as for the simple one-stage operational amplifier in order to appraise the feasibility of high frequency circuits and BBA circuits using V-NPN. Deep N-well CMOS technology and parasitic V-NPN are briefly described in Section II. The RF characteristics of V-NPN are presented in Section III. The RF mixer and simple one-stage operational amplifier using V-NPN are described in Sections IV and V, respectively. In Section VI, we propose two methods to increase the operating frequency of V-NPN for DCR, followed by the conclusion in Section VII.

II. Parasitic V-NPN in Deep N-Well CMOS

Nowadays, most of the state-of-the-art CMOS foundries provide the triple deep n-well technology [17]. The cross

Manuscript received January 6, 2004; revised July 7, 2004. This work is supported by MICROS (Micro Information and Communication Remote Object-Oriented Systems) Research Center.

The authors are with the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), MICROS Research Center, Daejeon 305-701, Korea (e-mail: nik@dimple.kaist.ac.kr; krlee@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/JSSC.2004.840982

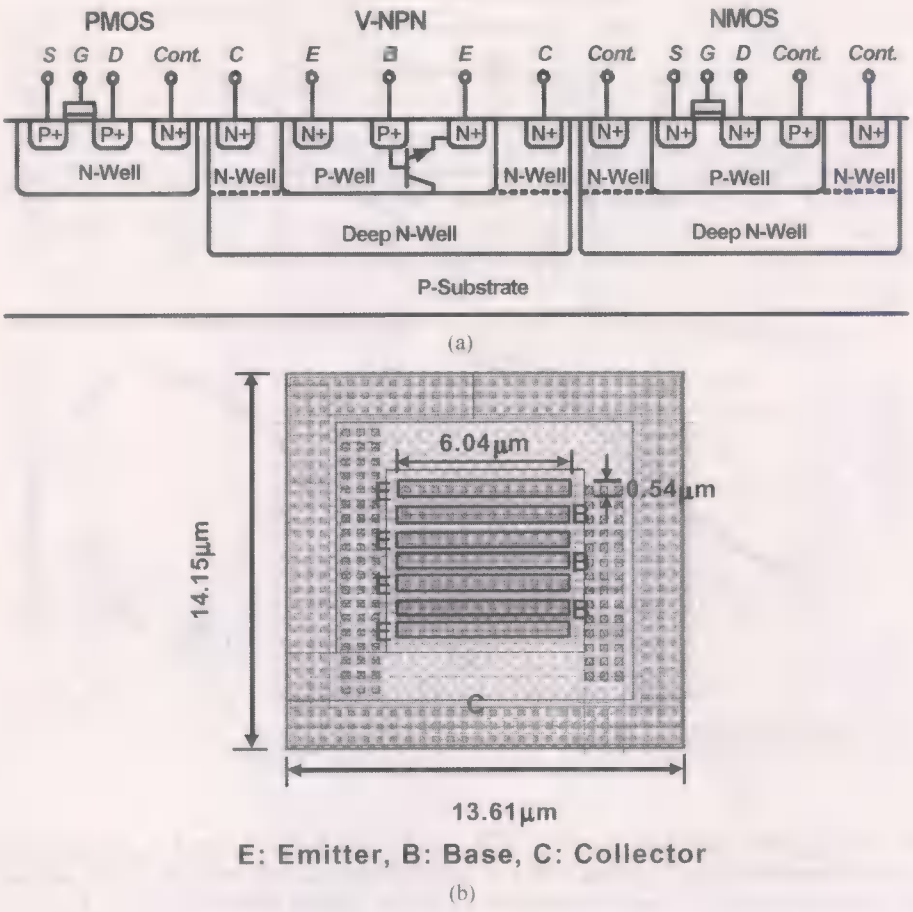


Fig. 1. (a) Cross sectional view of the deep n-well CMOS technology. (b) Layout for a V-NPN with four emitter fingers.

sectional view showing the well structure and various devices available from the deep n-well CMOS technology is presented in Fig. 1(a). The prime motivation for the deep n-well CMOS is that it is possible to apply different substrate bias to NMOS residing in other p-well so that we can adjust threshold voltages by electrical means, which is one of the most efficient ways to adaptively adjust power consumption. Moreover, this triple n-well CMOS technology, specifically deep n-well one, can provide excellent isolation against the substrate coupling noise among and between digital baseband logic circuits and RF and BBA circuits, which is especially important for integrating RF and baseband mixed mode circuits in a single chip. The deep n-well can completely isolate the p-well where NMOS is residing from the substrate coupling noise generated in other circuit blocks.

It should be noted that we can obtain high performance V-NPN free from this CMOS technology as shown in Fig. 1(a). It is composed of the n+ source-drain diffusion as the emitter, the p-well diffusion and p+ contact as the base, and deep n-well, n-well diffusion, and n+ contact as the collector. Deep n-well V-NPN provides not only lower collector resistance but also thinner p-base width, both of which can lead to high BJT performance. Note that the V-NPN differs from the previous parasitic substrate vertical BJT in that each collector is completely isolated. Since V-NPN has much better uniformity, reproducibility, device matching, driving capability, and more ideal BJT characteristics than the lateral one, we expect that the availability of this device can give us a great impact for mixed mode circuits such as DCR.

III. Electrical Characteristics of V-NPN

V-NPNs with various number of emitter fingers (1 to 5) were laid out and fabricated in deep n-well 0.18-μm 1-poly 6-metal CMOS foundry process. The area of each emitter finger is 0.54×6.04 μm^2. Fig. 1(b) shows the layout example for a V-NPN with four emitter fingers. The dc characteristics of this device were measured with an HP 4156 semiconductor parameter analyzer. Fig. 2(a) shows the collector current (I_C) versus collector voltage ($V_{\rm CE}$) curves measured with varying base current from 10 μA to 40 μA. 40 V of Early voltage, V_A, is obtained by extrapolating the active region of the curves in Fig. 2(a), which is much larger than MOSFET. DC current gain of 18, $BV_{\rm CBO}$ (collector-base breakdown voltage) of about 20 V and $BV_{\rm CEO}$ (collector-emitter breakdown voltage) of about 7 V are obtained. The Gummel plot is shown in Fig. 2(b). The curve of Fig. 2(c) shows that the current gain is almost constant over the wide range of collector current. At very low collector current, it depends on the collector current, indicating some nonideal base current characteristics. The maximum current gain of 18 is obtained at 22 μA of I_C. Note, however,

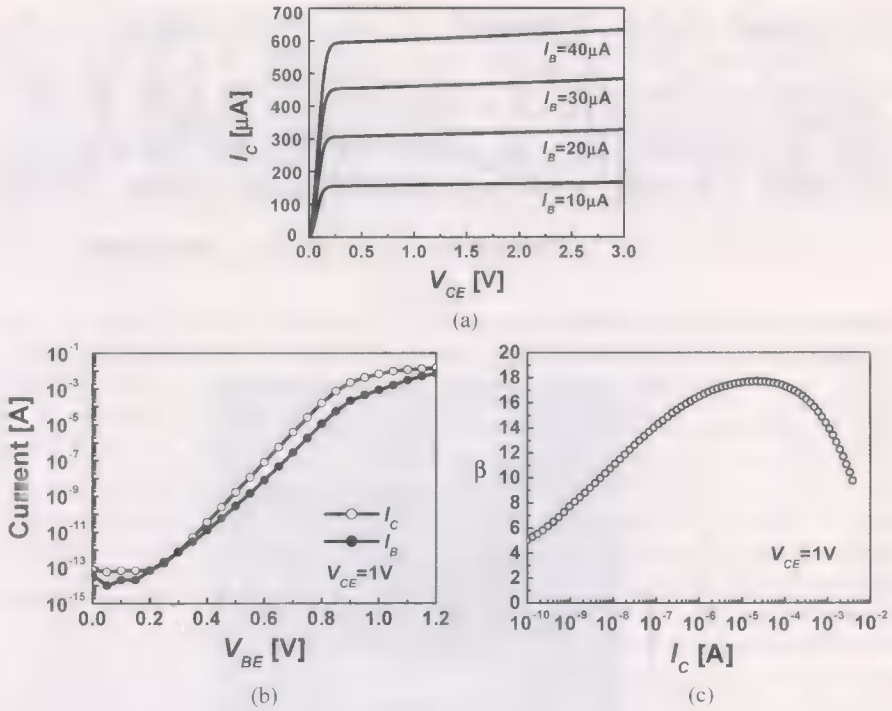


Fig. 2. DC characteristics of V-NPN with four emitter fingers: (a) collector current (I_B from 10 μA to 40 μA in steps of 10 μA); (b) Gummel plot; (c) β dependence on I_C.

that this dependence is much weaker than that in lateral NPN [13], showing much closer characteristics to an ideal BJT.

To see high-frequency characteristics of the V-NPN, S-parameters have been measured with HP 8510C network analyzer in the frequency range from 400 MHz to 6 GHz. The measured S-parameters were corrected for pad and interconnection parasitic contributions by means of open and short de-embedding patterns. The de-embedded spectra for the current gain $|h_{21}|^2$ and the MAG (maximum available gain) for V-NPN at 1.3 mA of collector bias current, are shown in Fig. 3(a). The unit current gain cutoff frequency f_t is 1.9 GHz and the maximum oscillation frequency $f_{\max}$ is 3.76 GHz. Fig. 3(b) plots the f_t and $f_{\max}$ versus I_C, showing peak f_t and $f_{\max}$ are obtained near 1 mA of I_C for this particular device. The unit current gain cutoff frequency is approximately given by

$$f_t = \left\{ 2\pi \left[\tau_F + \frac{kT(C_{\rm je} + C_{\rm jc})}{qI_C} \right] \right\}^{-1} \tag{1}$$

where τ_F is the forward charge-control time constant, $C_{\rm je}$ is the emitter-base junction capacitance, $C_{\rm jc}$ is the collector-base junction capacitance, k is Boltzmann's constant, T is absolute temperature, and q is the electronic charge [18]. Fig. 3(c) shows $1/f_t$ versus $1/I_C$ characteristics. From the y-intercept of this plot, we obtain τ_F of 85 ps. Assume that the value of τ_F is mainly dominated by the base transit time, τ_B, expressed as follows:

$$\tau_B \approx W_B^2/(2D_n) \tag{2}$$

where D_n is the diffusion constant for electrons, of which Boron is about 5.17[cm^{2}s^{-1}] at the given impurity concentration, and W_B is the base width [18]. The base width calculated from this is $W_B = 0.3$ μm [see Fig. 16(b)], which is very close to the process data, indicating f_t of this device is dominated by base transit time in vertical direction. Fig. 3(d) plots the peak f_t and $f_{\max}$ of V-NPNs with various number of emitter finger. Regardless of the number of emitter finger, f_t and $f_{\max}$ of V-NPN are about 2 GHz and 4 GHz, respectively. Also, this indicates that the high-frequency characteristics of V-NPN depend on not the parasitics due to the layout dependence but the base width.

Because V-NPN is a parasitic device, there is a concern for its uniformity. Therefore, we measured the parameters such as β, V_A, output resistance (r_o) and f_t on 30 samples of V-NPN with four emitter fingers fabricated in a same wafer under the same conditions as above. Fig. 4 plots the histograms of these parameters over samples. As shown in Fig. 4, V-NPN shows excellent uniformity within wafer of less than 3.7% for all the parameters studied in this paper.

On the other hand, the flicker noise of the V-NPN was measured with the low noise current preamplifier and spectrum analyzer. As shown in Fig. 5, the corner frequency of flicker noise for V-NPN is as low as 4 kHz at 0.5 mA of collector current. In contrast, the corner frequency of 20 μm/0.18 μm NMOS is about 3 MHz at the same current. As expected, the V-NPN has much better flicker noise performance, indicating the feasibility of mixer and BBA circuits fabrication with almost free $1/f$ noise.

IV. RF Mixer for DCR Using V-NPN

The output noise voltage of the down-conversion mixer using MOSFET for DCR can be calculated as $\overline{V_{\rm OUT}^2} = \overline{V_{\rm OUT\cdot NT}^2} + \overline{V_{\rm OUT\cdot NS}^2} + \overline{V_{\rm OUT\cdot R}^2}$ as shown in Fig. 6, where $V_{\rm OUT\cdot NT}$ is the noise generated in the transconductor, $\rm N_T$, $V_{\rm OUT\cdot NS}$ is that in

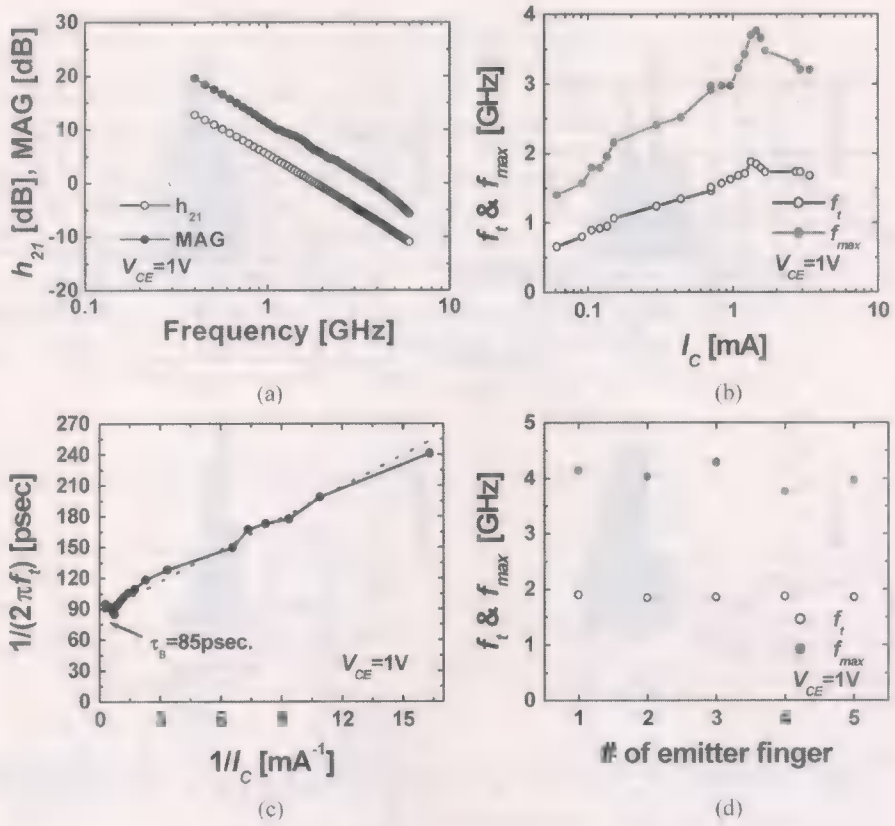


Fig. 3. RF characteristic of V-NPN: (a) the current gain $|h_{21}|^2$, and the maximum available gain (MAG); (b) cutoff frequency (f_t) and maximum oscillation frequency ($f_{\max}$) versus collector current (I_C); (c) $1/f_t$ versus $1/I_C$ plot showing base transit time of 85 ps; (d) peak f_t and $f_{\max}$ of V-NPNs with various number of emitter finger with unit finger area of 0.54 μm × 6.04 μm. All data are measured at $V_{CE} = 1$ V.

the switch, N_S, and $V_{OUT\cdot R}$ is that in the load resistor, R. Here, $\overline{V^2_{OUT\cdot NT}}$ can be expressed as

$$\overline{V^2_{OUT\cdot NT}} = 2 \times \left(4kT\gamma g_{d0\cdot NT} G_V^2 / g^2_{m\cdot NT}\right) \Delta f \quad (3)$$

where $g_{d0\cdot NT}$ is the drain conductance of N_T at $V_{DS} = 0$ V, γ represents the ratio of the value of thermal noise at any given drain bias to the value of thermal noise at $V_{DS} = 0$ V [19], $G_V = 2g_{m\cdot NT}R/\pi$ is the voltage gain of the mixer, $g_{m\cdot NT}$ is the transconductance of N_T, Δf is the bandwidth in hertz, and the factor 2 results from the two N_T's. The output noise voltage spectral density due to the switching pair and load resistor, $\overline{V^2_{OUT\cdot NS}}$ and $\overline{V^2_{OUT\cdot R}}$, can be expressed as

$$\overline{V^2_{OUT\cdot NS}} = 4 \times \left[4kT\gamma g_{d0\cdot NS} R^2 + K g^2_{m\cdot NS} R^2 / (C_{ox} W_{NS} L_{NS} f)\right] \Delta f. \quad (4)$$

and

$$\overline{V^2_{OUT\cdot R}} = 2 \times (4kTR)\Delta f \quad (5)$$

respectively. Here K is a process-dependant constant for $1/f$ noise (see Fig. 5), C_{ox} is the gate oxide capacitance per unit area, W_{NS} is the width of N_S, L_{NS} is the channel length of N_S, the factor 4 in (4) comes from the four N_S's, and the factor 2 in (5) comes from the two R's.

As shown in (4), the low-frequency noise is dominated by $1/f$ noise. Thus, we expect very small low-frequency noise in the mixer adopting V-NPN in the switching pair. To demonstrate this, we designed and fabricated a double-balanced RF mixer for DCR using V-NPN introduced in Section III, as shown in Fig. 7. Note, however, we still use NMOS (80 μm/0.18 μm) transconductors, because it provides higher linearity and gain with 1 mA of total mixer core current. The chip photograph is shown in Fig. 8. In order to minimize the parasitic capacitance C_{CS} between the collector and the substrate, the collectors of V-NPN switching transistor pair Q_1 and Q_3, and Q_2 and Q_4 were shared, respectively. The RF mixer was laid out as symmetrically as possible.

The measured conversion gain versus RF frequency is shown in Fig. 9. For the measurement, IF frequency is chosen at 1 MHz. When the RF frequency is over 2.4 GHz, the conversion gain decreases. It is very interesting to note that this mixer's 3-dB cutoff frequency is about 2.4 GHz, which is higher than the maximum f_t of 2 GHz. We believe that this is due to the frequency doubling effect of the differential circuits [20]. This fact is quite an encouraging result and is thought to be the characteristics of double-balanced mixer. Fig. 10 plots the IP_3 measurement results when two tones at 902.5 MHz and 903.5 MHz are mixed with LO frequency of 900 MHz and two tones at 2102.5 MHz and 2103.5 MHz are mixed with the LO frequency of 2100 MHz, respectively. IIP_3 is measured as -3.2 dBm and -5 dBm.

Fig. 11 presents the measured noise figure. As expected, the mixer has excellent low frequency noise performance, showing only thermal noise and almost $1/f$-noise-free characteristic. Therefore, the RF mixer using V-NPN switching transistors can be used even in very narrowband DCR such as for GSM.

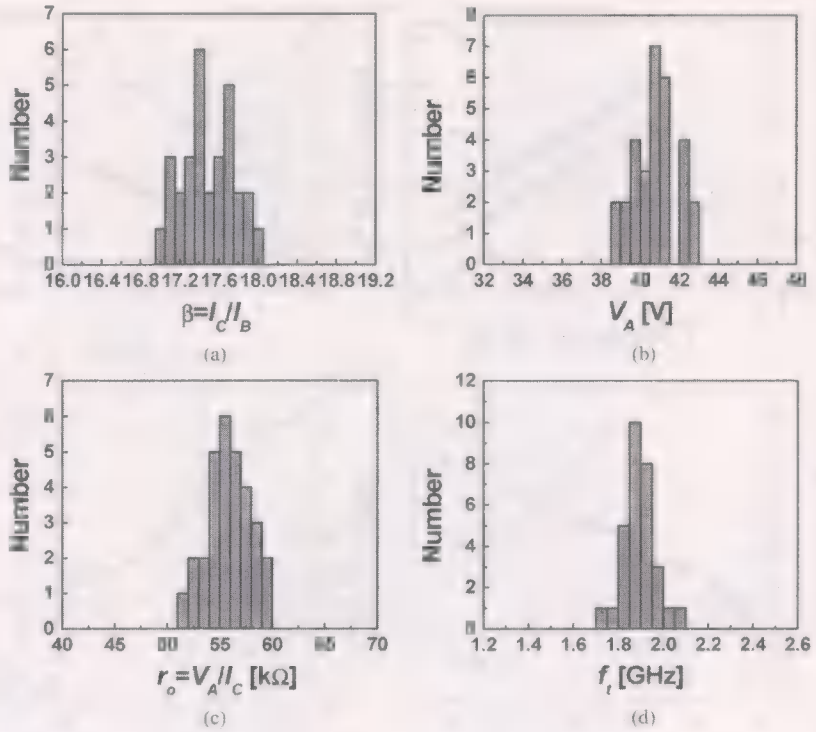


Fig. 4. Histograms of V-NPN parameters measured with 30 samples: (a) current gain (β); $M_\beta = 17.49, \sigma_\beta = 0.27$; (b) Early voltage ($V_A$); $M_V = 40.79$ V, $\sigma_V = 1.15$ V; (c) output resistance ($r_o = V_A/I_C, I_C = 0.73$ mA); $M_r = 56.17$ k$\Omega, \sigma_r = 2.06$ k Ω; (d) cutoff frequency (f_t); $M_f = 1.89$ GHz, $\sigma_f = 0.07$ GHz. All data are measured at $V_{CE} = 1$ V. (M: mean, σ: standard deviation).

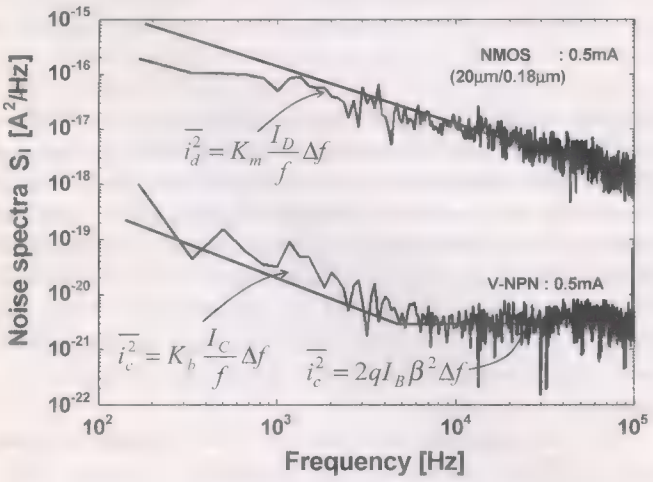


Fig. 5. Measured output noise spectra of V-NPN with four emitter fingers and NMOS of 20 μm/0.18 μm at 0.5 mA. The solid lines are $1/f$ noise models fitted with $K_m = 3 \times 10^{-10}$ and $K_b = 4 \times 10^{-14}$.

The output dc offset voltage of the mixer using V-NPN switching pair is shown in Fig. 12 measured as a function of LO input power, zero power limit of which is 0.6 mV. On the other hand, typical value for that of the mixer using NMOS switching transistors (aspect ratio; 50 μm/0.18 μm) is measured as 5–10 mV. This order of magnitude improvement is due to the much better device-to-device matching characteristic of V-NPN compared with NMOS device. Fig. 12 shows that the dc offset voltage increases as the LO input power and the LO frequency increase, as it should do because of the LO self-mixing.

Table I compares the performances of the V-NPN mixer against those of other published CMOS mixers. Clearly, we can obtain eminent noise figure and IIP_2 performance in the V-NPN mixer due to V-NPN characteristics such as low $1/f$ noise and good device-to-device matching. The parasitic V-NPN in deep n-well CMOS process can provide good enough mixer performance, opening a new horizon for low-cost CMOS DCR.

V. Operational Amplifier Using V-NPN

In addition to RF front-end, BBA circuits are also an important part in the wireless communication circuits. An operational amplifier is an essential part of BBA circuits such as active *RC* filter, programmable gain amplifier, etc. CMOS operational amplifiers (op amps) suffer from many problems such as large $1/f$ noise, large input offset voltage, and so forth. At low source impedance, the equivalent input noise voltage of one-stage CMOS op amp in Fig. 13(a) is expressed as [21]

$$\overline{V_{n,n}^2} = 2\left\{4kT(2/3)/g_{m1} + K_N/(C_{ox}W_1L_1f) + g_{m3}^2/g_{m1}^2[4kT(2/3)/g_{m3} + K_P/(C_{ox}W_3L_3f)]\right\}\Delta f. \quad (6)$$

The equivalent input noise voltage is mainly dominated by that of the differential NMOS input pair. As can be seen from (6), increasing the gate area of the input transistors can reduce the $1/f$ noise. However, its unavoidable penalties are greatly increased area and large input capacitances, both of which inevitably increase die size as well as the power consumption [14].

The alternative to large gate area of the NMOS input transistors is to adopt BJT in the input stage. To assess the feasibility of using V-NPN in BBA circuits, a simple one-stage differential op

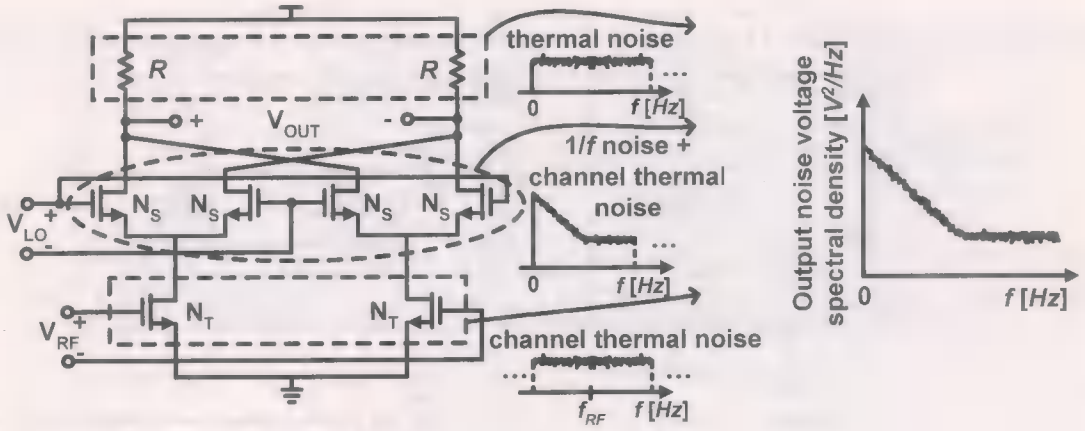


Fig. 6. The output noise voltage spectral density of double-balanced Gilbert mixer using MOSFET.

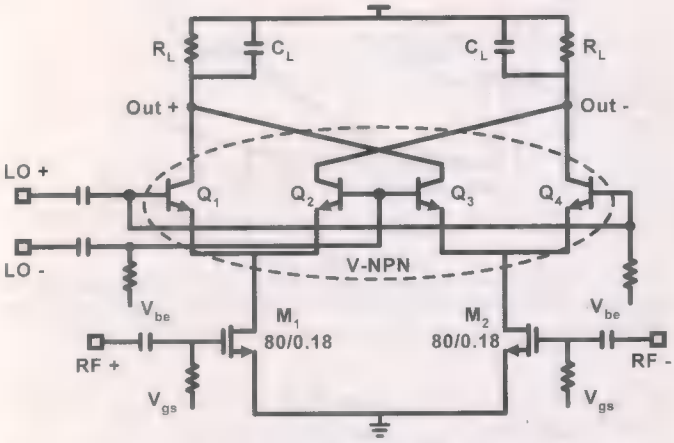


Fig. 7. Circuit schematic diagram of double-balanced Gilbert mixer using V-NPN.

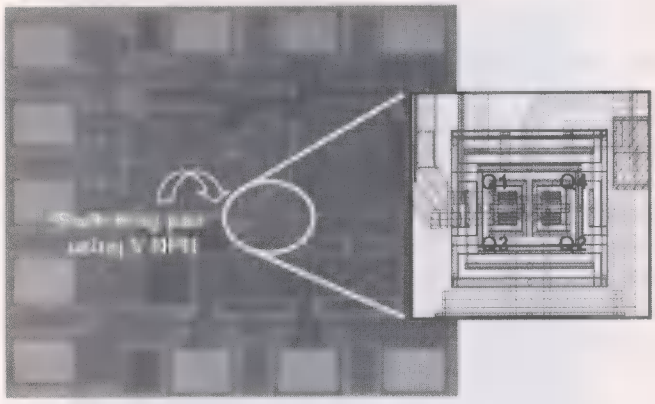


Fig. 8. Chip photograph of RF mixer using V-NPN switches.

amp has been designed, as shown in Fig. 13(b). The equivalent input noise voltage of one-stage V-NPN op amp in Fig. 13(b) is expressed as

$$\overline{V_{n\cdot v}^2} = 2\left\{4kTr_b + 4kT/(2g_{mv}) + g_{m3}^2/g_{mv}^2[4kT(2/3)/g_{m3} + K_P/(C_{ox}W_3L_3f)]\right\}\Delta f \quad (7)$$

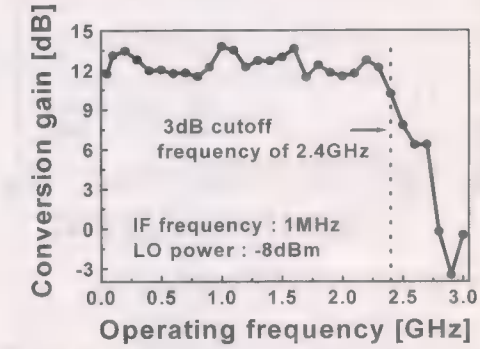


Fig. 9. Measured conversion gain versus RF frequency.

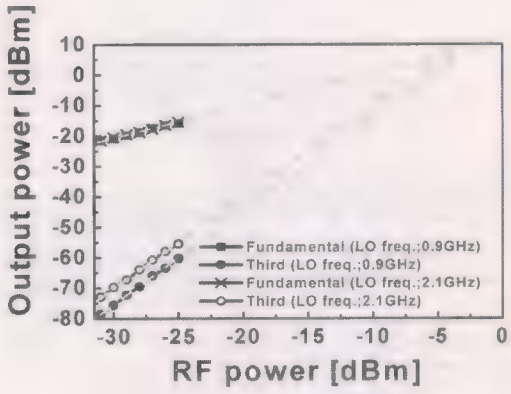


Fig. 10. IP_3 plot measured at LO input power of -8 dBm. The IIP_3 is -3.2 dBm and -5 dBm, respectively.

where r_b is the base resistance of Q_1. Because V-NPN has much larger transconductance $g_{mv}(= qI_C/kT)$, smaller $1/f$ noise than MOSFET, we expect much better noise performance through (7). Moreover, because the Early voltage V_A and the output resistance $r_{ov}(= V_A/I_C)$ are larger, much larger voltage gain $|A_V| \approx g_{mv}(r_{ov}//r_{o3})$ can be obtained at the same bias current. The only significant disadvantage of V-NPN op amp as compared to a CMOS one is the input bias current. The equivalent input noise current of a CMOS one is usually negligible due to very small input bias currents. However, the V-NPN op amp has a significant input noise current I_n generated by the base currents of the V-NPN input transistors.

TABLE I
MEASURED PERFORMANCE SUMMARIES OF RF MIXER USING V-NPN AND COMPARISON TO OTHER CMOS MIXERS ALREADY PUBLISHED

Operating frequency	0.9GHz (This work)	2.1GHz (This work)	[13]	0.9GHz [27]
Conversion gain	12dB	12dB	15dB	18dB
IIP_2	> 40dBm	> 30dBm	44dBm	30dBm
IIP_3	-3.2dBm	-5dBm	-8.2dBm	-4dBm
NF (DSB)	7.6dB	8.5dB	17.8dB	18dB(SSB)
Power consumption	0.9mA @1.8V	0.9mA @1.8V	0.73mA @ 3V	6mA @ 2.7V
Technology	0.18μm CMOS	0.18μm CMOS	0.35μm CMOS	0.35μm CMOS

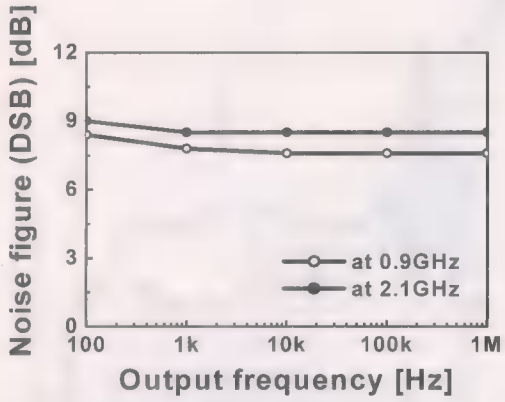


Fig. 11. Noise figure measured at 0.9 GHz and 2.1 GHz.

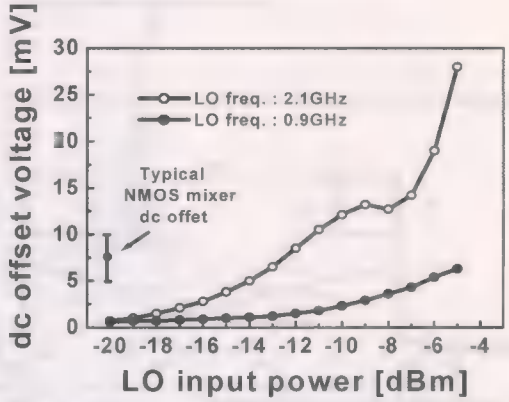


Fig. 12. The dc offset voltage of RF mixer using V-NPN switching pair versus input power level. (The data indicated by an error bar is the range of the dc offset measured from NMOS mixer fabricated using same CMOS technology).

On the other hand, the input offset voltage of the CMOS op amp, $V_{\mathrm{OS}\cdot n}$, and that for the V-NPN op amp, $V_{\mathrm{OS}\cdot v}$, can be approximated respectively as [22]

$$V_{\mathrm{OS}\cdot n} \approx V_{\mathrm{TH1}} - V_{\mathrm{TH2}} + (V_{\mathrm{TH3}} - V_{\mathrm{TH4}}) \times \sqrt{\frac{\mu_p (W/L)_P (1 + |\lambda_p V_{\mathrm{DSP}}|)}{\mu_n (W/L)_N (1 + \lambda_n V_{\mathrm{DSN}})}} + \frac{1}{2}\sqrt{\frac{I_M}{\mu_n C_{\mathrm{ox}} (W/L)_N (1 + \lambda_n V_{\mathrm{DSN}})}} \times \left(\frac{\Delta(W/L)_P}{(W/L)_P} - \frac{\Delta(W/L)_N}{(W/L)_N} \right) \quad (8)$$

and

$$V_{\mathrm{OS}\cdot v} \approx V_T \left(\frac{\Delta(W/L)_P}{(W/L)_P} + (V_{\mathrm{TH3}} - V_{\mathrm{TH4}}) \times \sqrt{\frac{\mu_p C_{\mathrm{ox}} (W/L)_P (1 + |\lambda_p V_{\mathrm{DSP}}|)}{I_M/4}} - \frac{\Delta I_S}{I_S} \right). \quad (9)$$

Here V_{TH} is the threshold voltage, $(W/L)_N = (W_1/L_1 + W_2/L_2)/2$ is the combined W/L of M_1 and M_2, $(W/L)_P = (W_3/L_3 + W_4/L_4)/2$ is that of M_3 and M_4, λ is the channel length modulation coefficient, V_{DSN} is the drain-source voltage of M_1 and M_2, V_{DSP} is the drain-source voltage of M_3 and M_4, $\Delta(W/L)_N = W_1/L_1 - W_2/L_2$, $\Delta(W/L)_P = W_3/L_3 - W_4/L_4$, μ_n is the mobility of electrons, μ_p is the mobility of holes, V_T is the thermal voltage, $I_S = (I_{S1} + I_{S2})/2$, $\Delta I_S = I_{S1} - I_{S2}$, I_{S1} is the scale current of Q_1, and I_{S2} is the scale current of Q_2. Note that (9) is derived here following similar procedure for (8). Because the effect of V_T in (9) can be scaled by $\Delta(W/L)_P/(W/L)_P$, it can be known that $V_{\mathrm{OS}\cdot v}$ would be much smaller than $V_{\mathrm{OS}\cdot n}$ from the (8) and (9).

The chip photograph of the fabricated V-NPN op amp is shown in Fig. 14. Table II summarizes the performance of CMOS op amp and V-NPN op amp. The V-NPN op amp has the voltage gain of 58 dB, equivalent input noise voltage (V_n) of 2.9 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ with $1/f$ corner frequency (f_n) of 1.9 kHz, and equivalent input noise current of 0.7 $\mathrm{pA}/\sqrt{\mathrm{Hz}}$ with f_n of 1.8 kHz. Especially, V-NPN op amp has two order of magnitude lower f_n and smaller V_n^2 than CMOS one at the same current. Furthermore, its input offset voltage is about 1 mV, which is much smaller than that in CMOS. The input base current of V-NPN differential pair is 1.54 μA, respectively. The input offset current between V-NPN differential pair is measured about 5 nA using HP4142 B. Since V-NPN device-to-device matching is excellent, the impact of input offset current is negligible.

VI. WAYS TO INCREASE OPERATING FREQUENCY OF V-NPN

As stated above, it is known that the RF mixer and operational amplifier using V-NPN are much robust against the low-frequency noise and mismatch, both of which are vital to DCR. For example, the utilization of V-NPN as shown in Fig. 15 makes high-performance CMOS DCR possible. Also, by combining V-NPN and MOSFET devices on the same chip, we can optimize the analog/digital circuits and maximize the tradeoff between speed and power. Therefore, V-NPN can give impact on the implementation of high-performance CMOS DCR as well as system-on-a-chip.

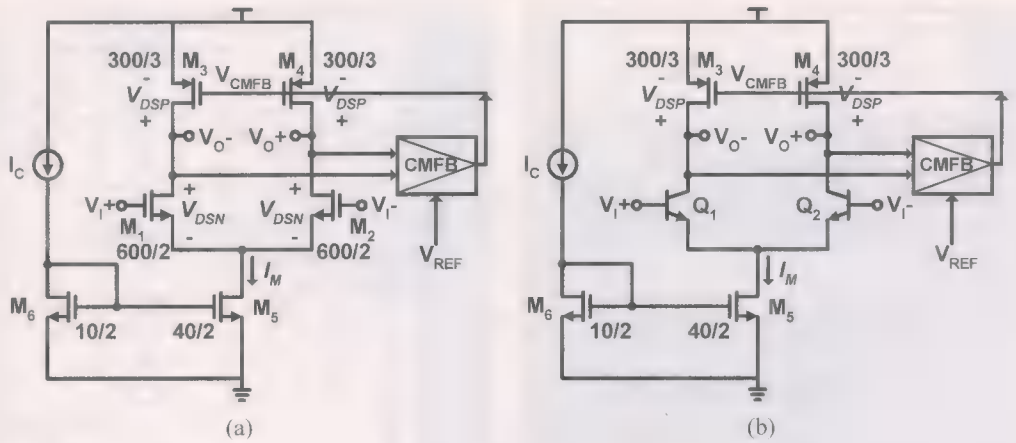


Fig. 13. Circuit schematic diagram of (a) one-stage CMOS operational amplifier and (b) one-stage V-NPN operational amplifier.

TABLE II
PERFORMANCE SUMMARIES OF CMOS OPERATIONAL AMPLIFIER AND V-NPN OPERATIONAL AMPLIFIER

	CMOS operational amplifier (simulation)	V-NPN operational amplifier
Voltage gain	49dB	58dB
V_n @ 1kHz	41.5 nV/$\sqrt{\text{Hz}}$	4.3 nV/$\sqrt{\text{Hz}}$
f_n (V_n)	310kHz	1.9kHz
V_n (at midband)	4.6 nV/$\sqrt{\text{Hz}}$	2.9 nV/$\sqrt{\text{Hz}}$
Input offset voltage	-	< 1mV
I_n @ 1kHz	-	1.1 pA/$\sqrt{\text{Hz}}$
f_n (I_n)	-	1.8kHz
I_n (at midband)	-	0.7 pA/$\sqrt{\text{Hz}}$
Input bias current	-	1.54μA
Input offset current	-	5nA
Power consumption	120μA @1.8V	128μA @1.8V

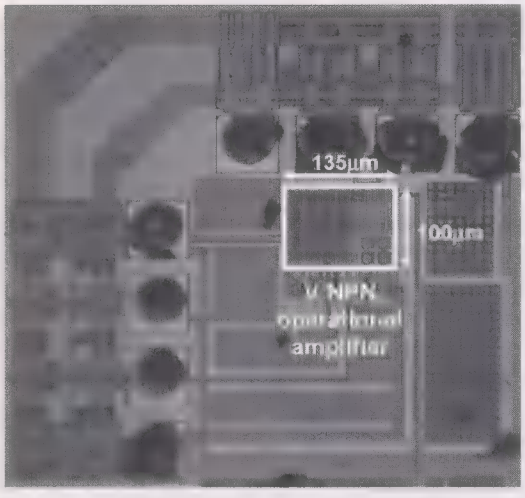


Fig. 14. Chip photograph of V-NPN operational amplifier.

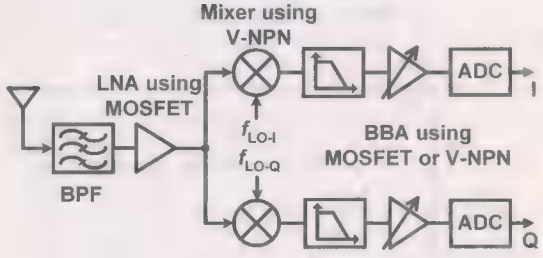


Fig. 15. The impact of V-NPN for single-chip radio.

However, the current V-NPN circuit has very limited RF performance because its f_t is an order of magnitude lower than that of MOSFET. Due to its low f_t, it is difficult to apply V-NPN to higher frequency circuits. In this paper, we propose two ways to increase its operating frequency. One is a simple fabrication process change and the other is a receiver architecture change.

Fig. 16 shows how thin base width can be obtained in two ways. One is to use a separate shallower p-well implant and the other is to use shallower deep n-well implant processes. To validate this simply, V-NPN with four emitter fingers was simulated using Athena and Atlas [23]. We followed the same process steps as in [24]. Fig. 16(a) shows the simulated cross view and Fig. 16(b) plots the two-dimensional (2-D) net doping profile of the V-NPN through the cutting-plane line A in Fig. 16(a). The f_t versus base width by keeping peak base doping constant at $5 \times 10^{17}/\text{cm}^3$ is shown in Fig. 17(a) before collector-to-emitter punchthrough at $V_{\text{CE}} = 1$ V. Fig. 17(b) shows how f_t of V-NPN can also be improved by changing deep n-well implantation energy before pinch-off at $V_{\text{CE}} = 1$ V. As can be seen, more

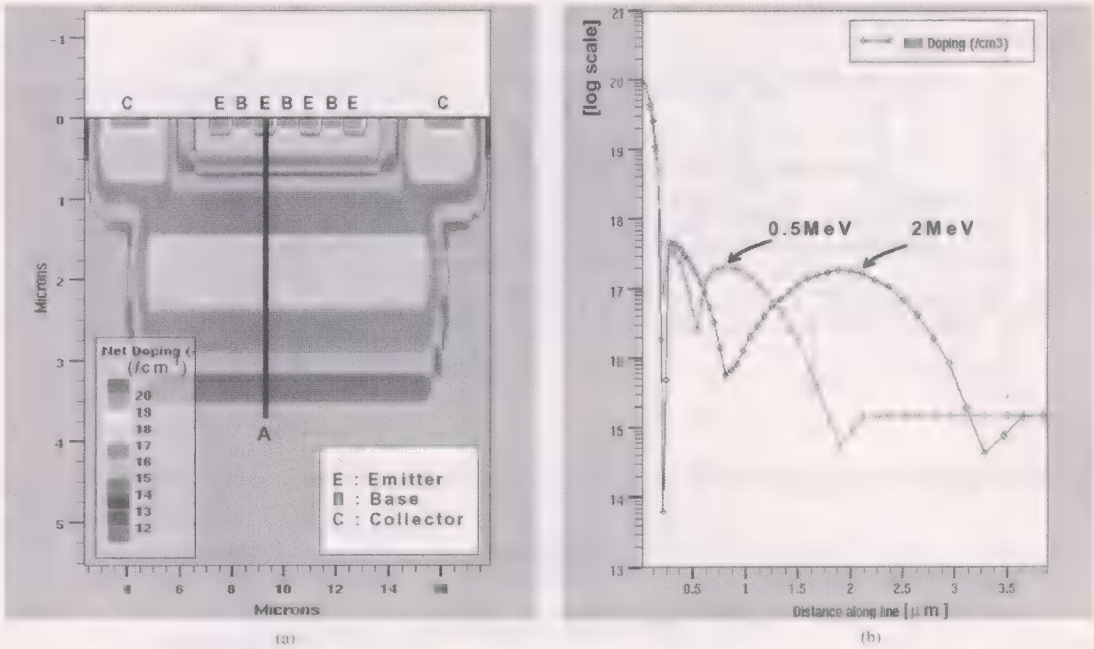
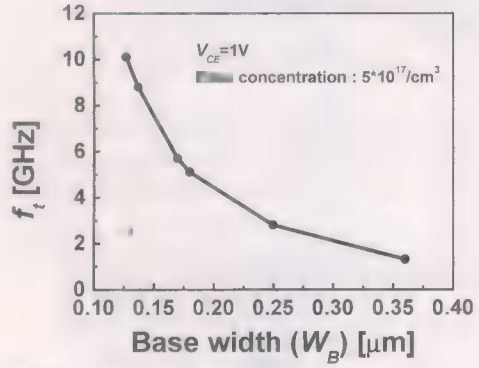
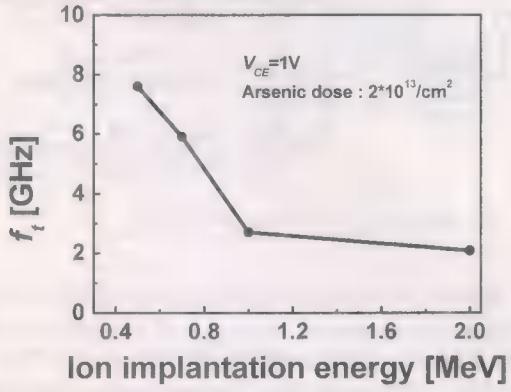


Fig. 16. (a) Simulated cross view and (b) 2-D net doping profile of V-NPN for deep n-well implantation dose of 2×10^{13} cm^2 with two different energy of 0.5 MeV and 2 MeV before punchthrough at $V_{CE} = 1$ V.



(a)



(b)

Fig. 17. (a) f_t versus the base width of V-NPN and (b) f_t versus deep n-well implantation energy.

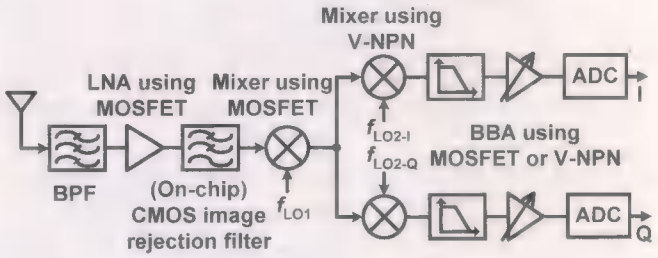


Fig. 18. Dual conversion receiver adopting V-NPN. Note that dual conversion high-IF receiver allows on-chip image rejection filter implementation in CMOS [28], [29].

than 10 GHz of f_t can be readily obtained with one additional process.

The second method is to change the receiver architecture, that is, to adopt the dual-conversion receiver [25] as shown in Fig. 18. The advantages of the dual-conversion receiver are as follows: no RF channel-select frequency synthesizer required, design flexibility (for example, giving gain at IF stages), less dc offset, weak LO pulling, and low LO leakage, compared with DCR. However, the dual-conversion receiver has disadvantages in which additional mixers require more power, noise, and distortion, image rejection filter augments the die area, and image rejection is limited by gain matching and LO deviation from quadrature [26]. Because the second mixer and following BBA circuits of the dual-conversion receiver process the baseband signal, the $1/f$ noise and dc offset characteristics of these blocks have a considerable influence on the baseband signal. Therefore, if the LNA and first mixer are implemented using MOSFET devices with high f_t and the second mixer and following BBA circuits are implemented with the combination of V-NPN and MOSFET, the operating frequency can greatly be extended exploiting all the advantages of V-NPN circuits. In the same way, this can be applied to the Weaver DCR [26] as in Fig. 19 that has the image rejection capability by the self-aligning image-rejection mixer. Therefore, the pertinent use of V-NPN and MOSFET in the dual-conversion receiver and Weaver DCR can extend the operating frequency of DCR with all the inherent advantages of V-NPN DCR.

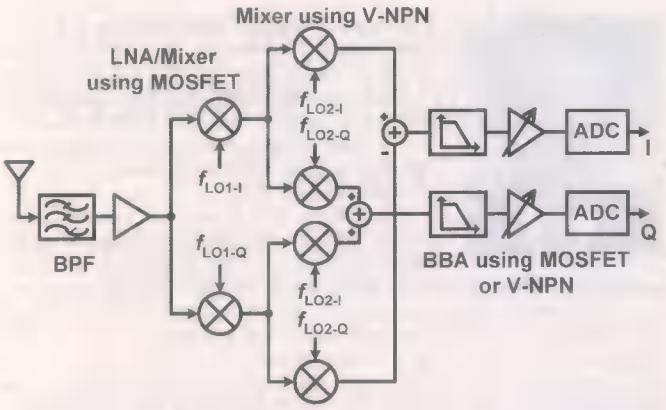


Fig. 19. Weaver DCR adopting V-NPN.

VII. Conclusion

We have presented the electrical characteristics of V-NPN available in deep n-well 0.18-μm CMOS technology. A double-balanced RF mixer using V-NPN shows almost free of $1/f$ noise as well as an order of magnitude smaller dc offset with other characteristics comparable with the CMOS one and 12 dB flat gain up to the frequency higher than the current cutoff frequency of the V-NPN transistor itself. The V-NPN operational amplifier for BBA circuits has higher voltage gain, better noise performance, and better matching than the CMOS one at the same current. These circuits using V-NPN can have great impact on the possibility of high-performance direct-conversion receiver implementation in CMOS technology. With further scaling of CMOS, and/or one additional base implant process step, and/or the adoption of the dual-conversion architectures and Weaver DCR, very high-performance DCR comparable to those obtained from pure bipolar or BiCMOS can be fabricated from low-cost CMOS technology.

Acknowledgment

The authors appreciate useful discussion with Dr. Y. J. Kim at Samsung Electronics and Dr. B. Kim at Integrant Technologies. The authors thank the reviewers for valuable comments and advice, and Dr. S. Hyun at ETRI and Dr. B. Kim at Integrant Technologies for their support.

References

[1] D. A. Rich, M. S. Carroll, M. R. Frei, T. G. Ivanov, M. Mastrapasqua, S. Moinian, A. S. Chen, C. A. King, E. Harris, J. D. Blauwe, H.-H. Vuong, V. Archer, and K. Ng, "BiCMOS technology for mixed-digital, analog, and RF applications," *IEEE Microwave Mag.*, vol. 3, no. 2, pp. 44–55, Jun. 2002.

[2] L. E. Larson, "Integrated circuit technology options for RFICs—Present status and future directions," *IEEE J. Solid-State Circuits*, vol. 33, no. 3, pp. 387–399, Mar. 1998.

[3] G. Chang, L. Jansson, K. Wang, J. Grilo, R. Montemayor, C. Hull, M. Lane, A. X. Estrada, M. Anderson, I. Galton, and S. V. Kishore, "A direct-conversion single-chip radio-modem for bluetooth," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, Feb. 2002, pp. 88–89.

[4] P. H. Woerlee, M. J. Knitel, R. van Langervelde, D. B. M. Klaassen, L. F. Tiemeijer, A. J. Scholten, and A. T. A. Z. Duijinhoven, "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, Aug. 2001.

[5] P. Choi, H. Park, I. Nam, K. Kang, Y. Ku, S. Shin, S. Park, T. W. Kim, H. Choi, S. Kim, S. Park, M. Kim, S. M. Park, and K. Lee, "An experimental coin-sized radio for extremely low-power WPAN (IEEE802.15.4) application at 2.4 GHz," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, Feb. 2003, pp. 92–93.

[6] T. Cho, E. Dukatz, M. Mack, D. MacNally, M. Marringa, S. Mehta, C. Nilson, L. Plouvier, and S. Rabii, "A single-chip CMOS direct-conversion transceiver for 900 MHz spread-spectrum digital cordless phones," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, Feb. 1999, pp. 228–229.

[7] B. E. Kim, S. Y. Kim, T. J. Lee, J. K. Lim, Y. J. Kim, M. S. Jeong, K. Kim, S. U. Kim, S. H. Park, and B. K. Ko, "A CMOS single-chip direct conversion satellite receiver for digital broadcasting system," in *Symp. VLSI Circuits Dig. Papers*, Honolulu, HI, Jun. 2002, pp. 238–241.

[8] B. Razavi, *RF Microelectronics*: Prentice Hall, 1998.

[9] E. A. Vittoz, "MOS transistors operated in the lateral bipolar mode and their application in CMOS technology," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 3, pp. 273–279, Jun. 1983.

[10] M. G. Degrauwe, O. N. Leuthold, E. A. Vittoz, H. J. Oguey, and A. Descombes, "CMOS voltage references using lateral bipolar transistors," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 6, pp. 1151–1157, Dec. 1985.

[11] T.-W. Pan and A. A. Abidi, "A 50-dB variable gain amplifier using parasitic bipolar transistors in CMOS," *IEEE J. Solid-State Circuits*, vol. 24, no. 4, pp. 951–961, Aug. 1989.

[12] S. Ye, K. Yano, and C. A. T. Salama, "A 1 V, 1.9 GHz mixer using a lateral bipolar transistor in CMOS," in *Proc. Int. Symp. Low Power Electronics and Design*, Aug. 2001, pp. 112–116.

[13] Z. Zhang and J. Lau, "A flicker-noise-free dc-offset-free harmonic mixer in a CMOS process," in *Proc. IEEE Radio and Wireless Conf.*, Waltham, MA, Aug. 2001, pp. 113–116.

[14] W. T. Holman and J. A. Connelly, "A compact low noise operational amplifier for 1.2 μm digital CMOS technology," *IEEE J. Solid-State Circuits*, vol. 30, no. 6, pp. 710–714, Jun. 1995.

[15] Y. P. Tsividis and R. W. Ulmer, "A CMOS voltage references," *IEEE J. Solid-State Circuits*, vol. 13, no. 6, pp. 774–778, Dec. 1978.

[16] I. Nam, Y. J. Kim, and K. Lee, "Low $1/f$ noise and DC offset RF mixer for direct conversion receiver using parasitic vertical NPN bipolar transistor in deep N-well CMOS technology," in *Symp. VLSI Circuits Dig. Papers*, Kyoto, Japan, Jun. 2003, pp. 223–226.

[17] *XQXQXQ*, 2000. TSMC Documentation no. T-018-MM-TM-002.

[18] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed. New York: Wiley, 1986.

[19] K. Han, H. Shin, and K. Lee, "Analytical drain thermal noise current model valid for deep submicron MOSFETS," *IEEE Trans. Electron Devices*, vol. 51, no. 2, pp. 261–269, Feb. 2004.

[20] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 4th ed. New York: Oxford, 1998.

[21] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York: McGraw-Hill, 2001.

[22] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. New York: Wiley, 1999.

[23] *ATHENA and ATLAS User's Manual*, Silvaco Int., 1996.

[24] J.-G. Su, H.-M. Hsu, S.-C. Wong, C.-Y. Chang, T.-Y. Huang, and J. Y.-C. Sun, "Improving the RF performance of 0.18-μm CMOS with deep n-well implantation," *IEEE Electron Device Lett.*, vol. 22, no. 10, pp. 481–483, Oct. 2001.

[25] D. Su, M. Zargari, P. Yue, S. Rabii, D. Weber, B. Kaczynski, S. Mehta, K. Singh, S. Mendis, and B. Wooley, "A 5 GHz CMOS transceiver for IEEE 802.11 a wireless LAN," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, Feb. 2002, pp. 92–93.

[26] J. C. Rudell, J.-J. Ou, T. B. Cho, G. Chien, F. Brianti, J. A. Weldon, and P. R. Gray, "A 1.9-GHz wide-band IF double conversion CMOS receiver for cordless telephone applications," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 2071–2088, Dec. 1997.

[27] D. Manstretta, R. Castello, and F. Svelto, "Low $1/f$ noise CMOS active mixers for direct conversion," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 49, no. 9, pp. 846–850, Sep. 2001.

[28] H. Samavati, H. R. Rategh, and T. H. Lee, "A 5 GHz CMOS wireless LAN receiver front end," *IEEE J. Solid-State Circuits*, vol. 35, no. 5, pp. 765–772, May 2000.

[29] M. H. Koroglu and P. E. Allen, "A 1.9 GHz image-reject front-end with automatic tuning in a 0.15 μm CMOS technology," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, Feb. 2003, pp. 264–265.

Ilku Nam (S'02) was born in Seoul, Korea, in 1975. He received the B.S. degree in electronics engineering from Yonsei University, Seoul, in 1999 and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2001. He is currently working toward the Ph.D. degree at KAIST.

Since 2000, he has participated in the development of low-power RF front-end circuits, low-power analog baseband circuits, and the wireless SOC for low-rate wireless personal area network (LR-WPAN). His research interests include CMOS RF/analog IC and RF system design for wireless communication, and interfaces among RF, modem, and MAC layer.

Kwyro Lee (M'80–SM'90) received the B.S. degree in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and the M.S. and Ph.D. degrees from the University of Minnesota, Minneapolis, in 1981 and 1983, respectively, where he did many pioneering works for characterization and modeling of AlGaAs/GaAs heterojunction field effect transistor.

From 1983 to 1986, he worked as an Engineering General Manager with GoldStar Semiconductor Inc., Korea, responsible for the development of the first polysilicon CMOS products in Korea. He joined the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1987 as an Assistant Professor in the Development of Electrical Engineering, where he is now a Professor. He has more than 150 publications in major international journals and conferences. He is the principal author of the book *Semiconductor Device Modeling for VLSI* (Prentice Hall, 1993) and one of the co-developers of AIM-SPICE, the world's first SPICE run under Windows.

Dr. Lee is a Life Member of the Korean Institute of Electrical and Communications Engineers. From 1990 to 1996, he served as the Conference Co-Chair of the International Semiconductor Device Research Symposium, Charlottesville, VA. From 1998 to 2000, he served as the KAIST Dean of Research Affairs and the Dean of Institute Development and Cooperation. At the same time, he also served as the Chairman of the IEEE Korea Electron Device Chapter and is currently serving as the elected member of EDS AdCom. Since 1997, he has been the Director of the MICROS.

Highly Integrated Direct Conversion Receiver for GSM/GPRS/EDGE With On-Chip 84-dB Dynamic Range Continuous-Time ΣΔ ADC

Yann Le Guillou, Olivier Gaborieau, Patrice Gamand, Martin Isberg, Peter Jakobsson, Lars Jonsson, David Le Déaut, Hervé Marie, Sven Mattisson, Laurent Monge, Torbjörn Olsson, Sébastien Prouet, and Tobias Tired

***Abstract*—This paper describes a highly digitized direct conversion receiver of a single-chip quadruple-band RF transceiver that meets GSM/GPRS and EDGE requirements. The chip uses an advanced 0.25-μm BiCMOS technology. The I and Q on-chip fifth-order single-bit continuous-time sigma-delta ($\Sigma\Delta$) ADC has 84-dB dynamic range over a total bandwidth of $\pm$135 kHz for an active area of 0.4 mm^2. Hence, most of the channel filtering is realized in a CMOS IC where digital processing is achieved at a lower cost. The systematic analysis of dc offset at each stage of the design enables to perform the dc offset cancellation loop in the digital domain as well. The receiver operates at 2.7 V with a current consumption of 75 mA. A first-order substrate coupling analysis enables to optimize the floor plan strategy. As a result, the receiver has an area of 1.8 mm^2.**

***Index Terms*—Analog-to-digital conversion, BiCMOS, continuous time, dc offset, direct conversion, EDGE, front-end, GPRS, GSM, IIP2, low-noise amplifier (LNA), mixer, self-mixing, sigma-delta ($\Sigma\Delta$).**

I. Introduction

THE global system for mobile communication (GSM) launched the second-generation system (2G) for cellular communication on a worldwide market. Today, the trend is to increase the number of data applications and the data rates. Enhanced data rates for GSM evolution (EDGE)—a 2.75G system—triples the GSM data rate going from a Gaussian minimum shift keying (GMSK) with 1 bit per symbol to an 8-level phase shift keying constellation (8-PSK) with 3 bits per symbol. It uses the GSM infrastructure and has the same symbol rate of 270 kS/s. To keep 2.75G system solutions cost-effective, the bill of materials (BOM) must be reduced as well as the power consumption. In this perspective, a direct-conversion receiver (DCR) is a very attractive architecture [1]–[4]. It eliminates the need for both IF and image reject filtering and requires only a single oscillator (LO) as illustrated in Fig. 1. Using a high dynamic range ADC, analog gain control (AGC) can significantly be reduced and most of the selectivity can be achieved in the digital baseband processor. Integrating the high dynamic range ADC on the RF-IC, the CMOS baseband becomes purely digital. It can then take advantage of CMOS process shrinking to reduce the overall power consumption and cost over the generations.

This work presents a DCR with an on-chip 13.5-bit resolution $\Sigma\Delta$ ADC over a bandwidth of $\pm$135 kHz. Section II focuses on the DCR design and techniques used to address the well-known weakness of DCR such as LO leakage, the self mixing, the finite second-order intercept point (IIP2), etc. [4], [5]. A brief description of the 0.25-μm BiCMOS technology associated with a first-order substrate coupling analysis is provided in Section III. Experimental results obtained from silicon implementation are presented in Section IV. Finally, in Section V, conclusions are drawn.

Manuscript received January 20, 2004; revised May 28, 2004.

Y. Le Guillou and P. Gamand are with Philips Semiconductors, 14079 Caen, France, and also with the Laboratoire de Microélectronique ensIcaen Philips (LaMIP), 14079 Caen, France (e-mail: yann.le.guillou@philips.com).

O. Gaborieau, D. L. Déaut, H. Marie, L. Monge, and S. Prouet are with Philips Semiconductors, 14079 Caen, France.

M. Isberg, P. Jakobsson, L. Jonsson, S. Mattisson, T. Olsson, and T. Tired are with Ericsson Mobile Platforms, SE-221 83 Lund, Sweden.

Digital Object Identifier 10.1109/JSSC.2004.841036

II. Circuit Design

The quad-band DCR is shown in Fig. 2. The low-band (LB) term is used for the GSM900 and GSM850 systems (880–960 MHz) and the high-band (HB) is for the DCS1800 and PCS1900 systems (1805–1990 MHz).

The low-noise amplifiers (LNAs) consist of four differential transconductors recombined through a cascode stage into one common resistive load for each band. The RF outputs of the LNAs are AC coupled to the in-phase (I) and quadrature-phase (Q) mixers so that the low-frequency distortion generated by the second-order nonlinearities in the LNA is blocked to prevent leakage through the mixer. The multiplier cells of the mixers use 1/2 or 1/4 sub-harmonic LO signal when high-band or low-band is selected, respectively [6], [7]. This LO configuration ensures sufficient frequency separation between the VCO frequency and the largest received blockers and their harmonics. It avoids VCO pulling and the associated LO phase noise degradation that would degrade the sensitivity performance in presence of interferers. The baseband chain includes a third-order low-pass filter that prevents the interferers from saturating the high dynamic range 1-bit continuous-time fifth-order $\Sigma\Delta$ ADC. The bit-stream coming from the ADC drives a low-voltage slew-rate controlled digital output buffer.

A. Low-Noise Amplifier (LNA)

Usually to achieve the best compromise between gain, linearity, noise, and input matching, emitter degeneration is provided by an inductance [8]. The requirement to combine extremely low noise figure (NF) LNA in a small area tends to relax

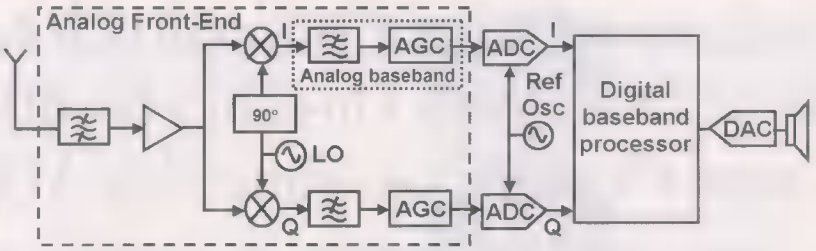


Fig. 1. Direct-conversion receiver architecture.

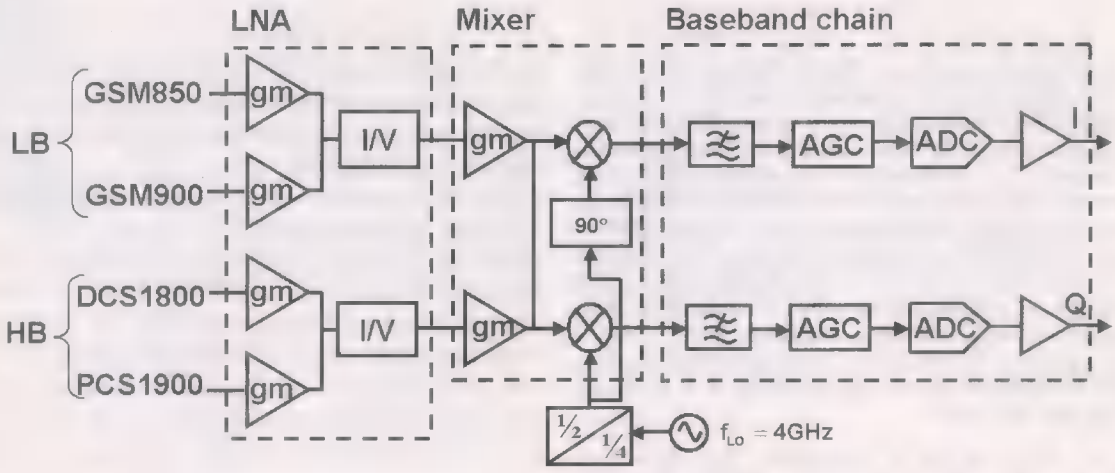


Fig. 2. Block diagram of the implemented quad-band DCR.

the inductive degeneration. Hence, the differential transconductance of each LNA is partly degenerated with an inductor and partly with an ac shunt feedback between the input and the output (see Fig. 3) [9]. However, in the ac shunt feedback configuration, the parallel input impedance depends on the feedback impedance Z_F and the resistive load R_L Then, to decrease the influence of the resistive load without increasing the area of the LNA, the parallel input impedance of 150 Ω is achieved by half an ac shunt feedback and by half an inductive degeneration feedback. After a parasitic extraction, the simulated return loss of this LNA is better than -22 dB in all bands. The NF is 2.2 dB whereas the gain and the ICP1 have been simulated respectively at 25 dB and -21 dBm for a current consumption of 8.7 mA.

B. Mixer

The I/Q direct conversion mixers are double-balanced Gilbert-type mixer topology as shown in Fig. 4. This topology provides inherently high IIP2 [4], [6], [7], [10], [11]. The resistive 100-Ω degeneration (R_{EE}) and the 7.8-mA current consumption has been chosen to trade off NF, IIP3, and input impedance. Since the $1/f$ noise spectrum of the mixers falls on top of the desired signal at baseband, only small NPN bipolar transistor with 40 GHz f_T were used in the mixers design to reduce the effect of flicker noise at the mixer output [11]. Transistor Q1 (Q2) drives Q5–Q6 (Q3–Q4) and Q9–Q10 (Q7–Q8) switch core transistors of BBI and BBQ, respectively. When properly scaled, the I_1, I_2, I_3, and I_4 current matching rely on the R_D resistors area while the switch core Q3–Q10 transistors can remain small. As a result, the parasitic capacitors are small. Thus, the LO transitions are sharp and the random modulation

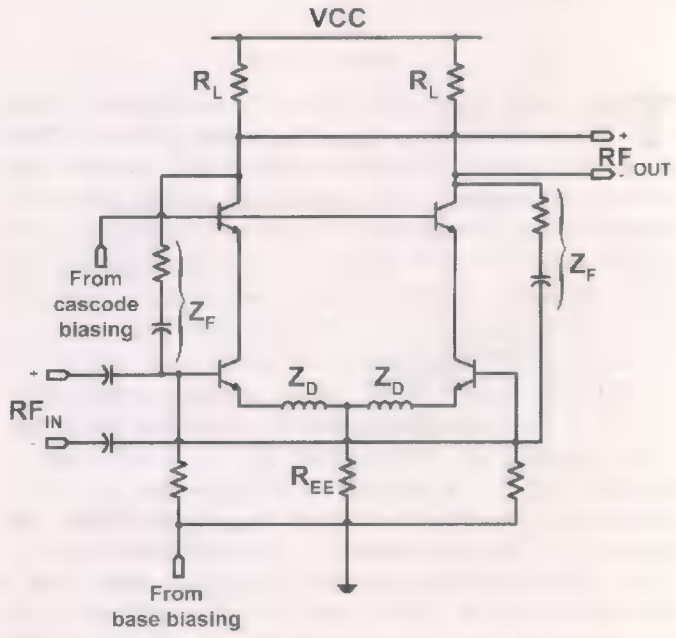


Fig. 3. Circuit diagram of the LNA.

of the switching time instants are small. Consequently, the flicker noise at the mixer output is minimized [11].

R_L and R_D are boron-doped polysilicon resistors. This type of resistors has good matching and flicker noise performance [12] as well as a precise $1/f$ noise modeling derived from experimental measurement.

A common centroid structure for the mixer-core and the mixer load R_L is required to compensate for thermal gradient

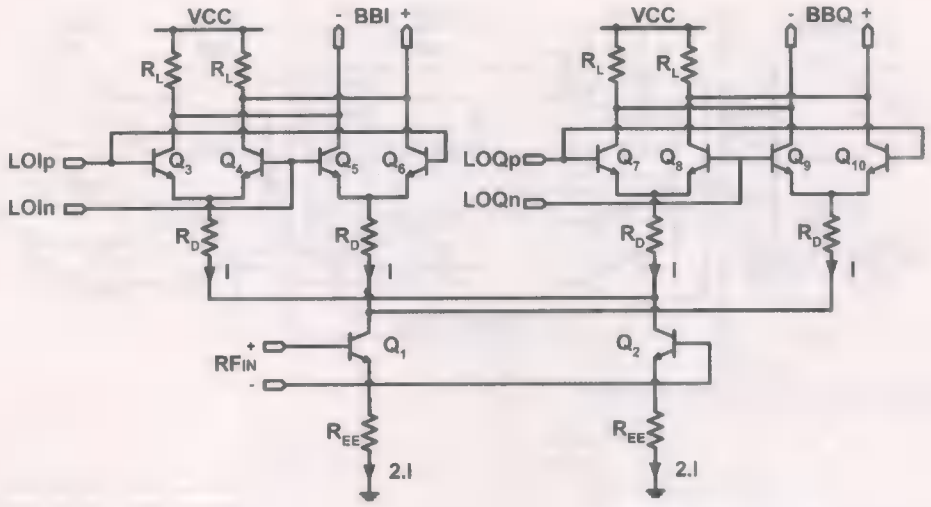


Fig. 4. Circuit diagram of the mixer.

effect and component mismatch that could result in gain and phase imbalances. The RF and LO lines crossing are perpendicular to reduce RF self-mixing.

C. LO Divider and Buffer

The VCO is running at two (HB) or four (LB) times the RF frequency. It prevents from VCO pulling [7]. The required LO dividers by 2 or 4 generate the I and Q quadrature.

As illustrated in Fig. 5, the fast divider by two uses a classical differential bipolar structure composed of master and slave latches. The 50-Ω R_D resistance is a compromise between I and Q matching, noise, and linearity requirement. Extensive Monte Carlo simulations after a parasitic extraction has enabled to fix the tail current (4I) at 4 mA to achieve an I/Q phase error lower than 1°. Typically, the LO signal edges slope is as sharp as 6 GV/s. It enables to minimize the flicker noise at the mixer output [11].

The LO signal path is 90° shifted from the RF signal path (see Fig. 13) to minimize the magnetic coupling between LNA and LO circuits.

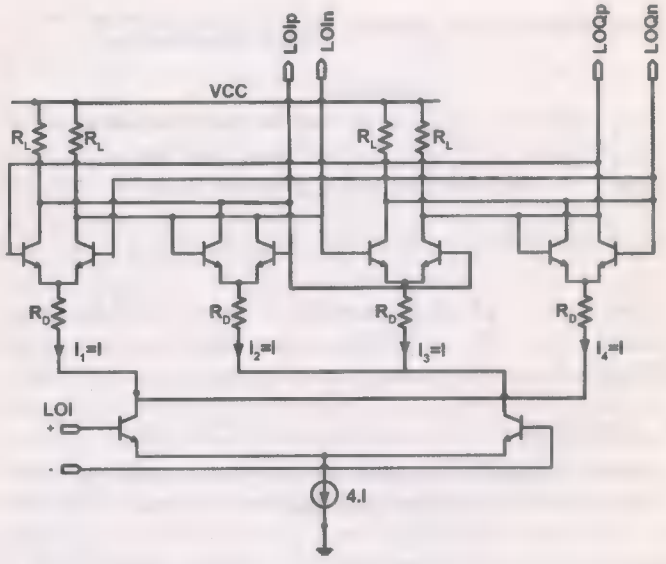


Fig. 5. Circuit diagram of the LO divider.

D. Baseband Filter Circuit

After downconversion and prior to digitization by the ADC, the baseband filter (BBF) completes the receiver chain. The BBF circuit enables the reduction of the dynamic range requirement on the ADC, through two mechanisms: first, by amplifying the wanted signal above the noise floor of the ADC, and second, by filtering the undesired signals—adjacent channels, blockers—so that they do not overload the ADC. As shown in Fig. 6, a third-order filter is sufficient to attenuate the worst blocker case, which is at 3 MHz. Hence, most of the channel filtering is performed in the digital baseband processor.

A first real pole is conveniently realized at the mixer output: its location early in the receiver chain alleviates the IP2 and IP3 requirements of following stages. The impedance is lower at the first stage of the baseband filter and scaled through the path to optimize the noise and the die area. As a result, the Sallen & Key stage, which is a complex pole filter, is introduced after 17 dB of gain in BB1. The global BBF amplification is 25 dB. Consequently, a 5-mV dc offset at the mixer output will result in 89-mV dc offset at BBF output. Hence, provided that the dc offset does not overload the ADC, the dc offset cancellation can be fully achieved in the digital baseband processor.

The unity gain buffer of the Sallen & Key stage, BB2, is introduced in the feedback path [13]. If located in the forward path, the buffer output impedance, together with feedback capacitor would build a parasitic zero, thus enlarging the out of band gain. The next amplifier, BB3, provides some gain trimming, and the final one, BB4, is designed to interface with the ADC.

The nominal 3-dB bandwidth of the BBF circuit is 208 kHz while the EDGE requirement is 135 kHz. This allowed more than 35% spread for process and temperature variations without corrupting the EDGE requirement. In addition, the group delay variation is below 0.18 μs even when the 3-dB bandwidth is at 135 kHz. Consequently, the BBF circuit does not need additional tuning circuitry to compensate for process and temperature variations.

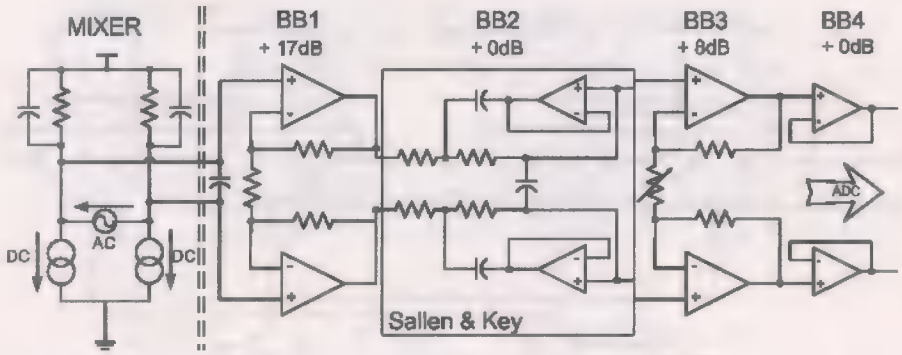


Fig. 6. Block diagram of one channel of the baseband filter (BBF).

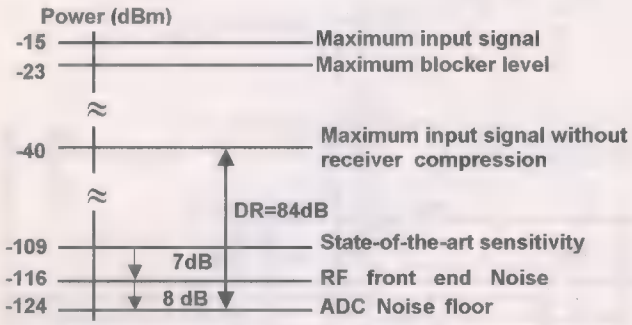


Fig. 7. ADC dynamic range requirement.

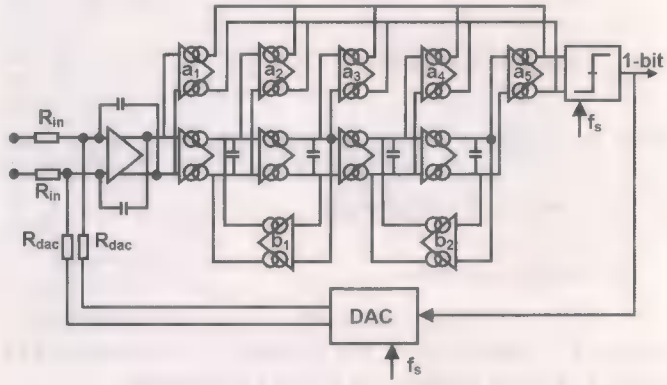


Fig. 8. Block diagram of the fifth-order CT $\Sigma\Delta$ modulator.

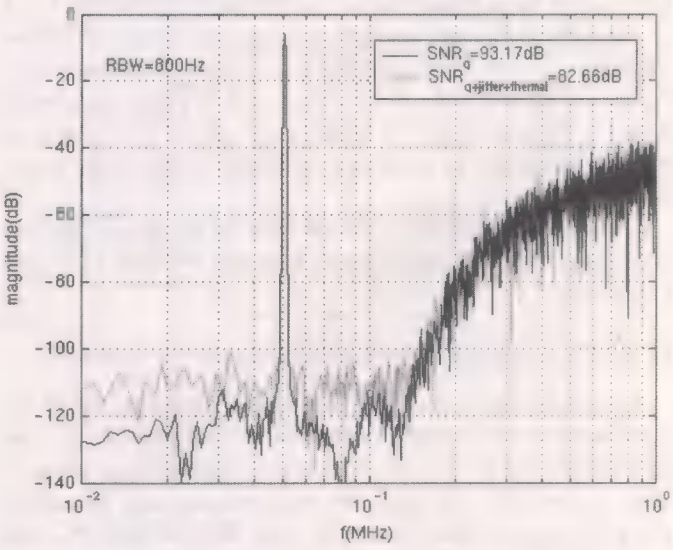


Fig. 9. Simulated output spectrum of the fifth-order CT $\Sigma\Delta$ modulator.

The total 2.5-nF capacitance for I and Q BBF has been stacked on the BBF active part to save area. As a result, the BBF is 0.5 mm^2 for a current consumption of 13 mA.

E. Fifth-Order Continuous-Time $\Sigma\Delta$ ADC Circuit

GSM/GPRS/EDGE requires the reception of signals between -104 dBm and -15 dBm [14]. The state-of-the-art sensitivity of -109 dBm is the target. The specified 10^{-3} bit-error rate (BER) requires a 7-dB signal-to-noise ratio (SNR) and thus leads to the system noise floor of -116 dBm. The ADC should not be the dominant noise source for a power efficient implementation. As illustrated in Fig. 7, its noise floor is 8 dB below the analog front-end's. If the out-of-band interferers are filtered so that they do not overload the ADC, the dynamic range requirement is then reduced to 84 dB and can be handled with a fifth-order $\Sigma\Delta$ ADC [15]. A low-pass continuous-time (CT) $\Sigma\Delta$ ADC is desirable since it enables a low-power implementation without the need for an anti-aliasing filter [16]–[19].

Fig. 8 shows the block diagram of the implemented ADC derived from [16]. The fifth-order loop filter has two complex conjugate poles introduced by the local feedback coefficients b_1 and b_2. They appear as notches in the shaped quantization noise (see Fig. 9). One of the notches is located at 78-kHz offset frequency. The other one is at the edge of the signal band. The feedforward coefficients a_i provide first-order roll-off at open-loop unity gain for stability reasons. Large signal stability is achieved by clipping the output integrator starting at the fifth integrator [18]. The input stage of the ADC consists of an operational transconductance amplifier in an integrating feedback configuration. The rest of the loop filter is implemented by means of transconductor-C (G_m-C) integrators for low-power reasons. The 1-bit feedback DAC is inherently linear. It switches resistors between positive and negative reference voltages derived from an on-chip bandgap reference circuit. A return-to-zero (RTZ) coding scheme is used to minimize the inter-symbol interference [20]. The biasing technique used for the design of temperature-insensitive g_m-C integrators avoids the need for tuning circuitry [19].

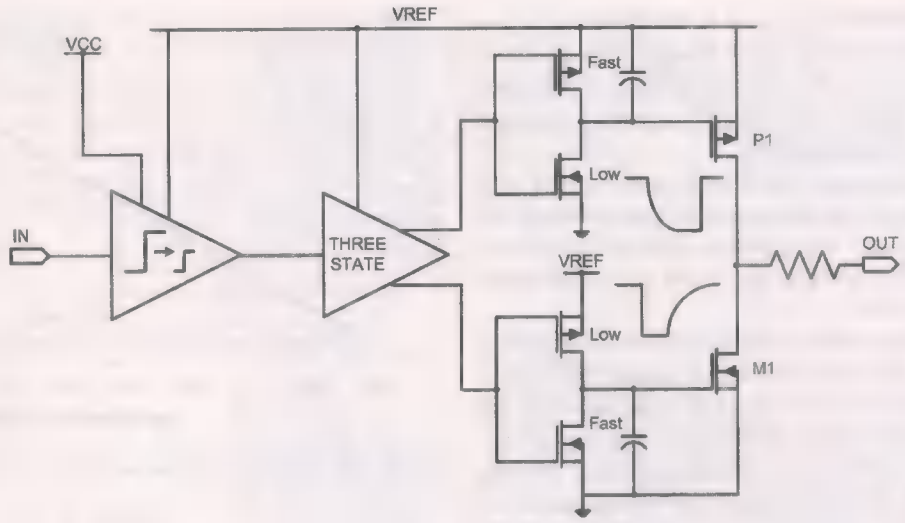


Fig. 10. Block diagram of the slew-rate limited output buffer.

As illustrated in Fig. 9, the simulated quantization noise is 93 dB below the maximum input signal at the modulator input, which is -3 dB compared to the overload level (-3 dBFs). This enables ± 150-mV dc offset at modulator input before overloading. Therefore, the dc offset cancellation can be fully achieved in the digital domain. The resistive DAC and the input stage of the ADC limit the SNR to 84 dB. This SNR is further degraded by 1.5 dB when the jitter of the 13-MHz clock is 5 ps_{rms}.

F. Slew-Rate Limited Output Buffer

The digital output buffers are used to drive the output pins with the 13-MHz bit-stream signals coming out of the ADC. Note that having a single-bit ADC allows to limit the required number of buffers to $3(I + Q + \text{clock})$ and thus to save power consumption and silicon area.

The digital output signal is shaped so as to limit its harmonics levels that might couple with input RF signal. For this purpose, a slew-rate limited buffer fed with a 1.5-V internally regulated V_{REF} is designed according to Fig. 10. The voltage regulation is made thanks to a classical series regulator together with a 100-pF MIM decoupling capacitor integrated over the entire block. The slew-rate limited buffer is made of two inverters in parallel feeding the power output PMOS and NMOS transistors with the digital signal. These inverters allow to put the buffer in tri-state mode by forcing the gates of the output PMOS and NMOS to V_{CC} and GND, respectively. They are sized to drive the output transistors in such a way to avoid direct current feedthrough from the output PMOS (NMOS) to the output NMOS (PMOS) during transitions. Hence, all the current is delivered to the load. A capacitor is added to the inverters output to help slowing down the current steered from V_{REF} (GND) through the output PMOS (NMOS). This helps limiting the high frequency harmonic levels on the output signal and avoiding large voltage spikes on the supply and ground rails. Finally, the 13-MHz clock frequency is slow enough to avoid high electromagnetic coupling with any close bond wires connected to sensitive blocks.

III. Process Implementation and Floorplan Strategy

A. Process Implementation

The quad-band receiver (as part of a fully integrated transceiver) has been fabricated in the RF 0.25-μm BiCMOS mature technology [21]. This technology features 40-GHz f_T and 90-GHz $f_{\max}$ NPN devices combined with high-quality passives and has been optimized for high frequency, low noise, and low supply current applications. Special effort has been put on the quality of passive components such as matching, quality factor, and deep trench isolation (DTI).

Low current consumption is achieved by optimizing the f_T versus the collector current and by using deep trench technique to reduce collector–substrate capacitance. For instance, this provides only 150-$\mu\text{A}/\mu\text{m}^2$ current density for 25-GHz f_T and less than 9-fF collector–substrate capacitance for a 0.4×20 μm^2 device.

Diffused and polysilicon resistors with less than $0.6\%{\cdot}\mu$m and $2.8\%{\cdot}\mu$m respective matching performance are particularly adequate for architectures where dc offset as well as I and Q mismatch need to be minimized.

Low k dielectric, thick metal, and DTI allow on-chip inductor Q as high as 20 at 2 GHz for a 1.5-nH coil.

The backend of this BiCMOS technology has been optimized to allow high routing density. It includes an embedded high-density 5-nF/mm^2 MIM capacitor built close to the top metal levels, which ensures very low parasitic elements to the substrate.

B. Floor Plan Strategy

A high level of function integration increases the sensitivity of the circuit to crosstalk. Sources of interferences are related to digital to analog coupling, electromagnetic (EM) coupling between inductors, routing traces coupling, interconnections, and signal injection through the substrate. Several tools exist to estimate these effects but they require a huge computation time and therefore do not allow fast design/layout iterations.

The methodology we have put in place is based on a simple model that "simulates" point-to-point effects due to EM or substrate coupling as a function of the distance, substrate resistivity,

and frequency. The model consists of a "black box" $H(j\omega)$, which is connected between two circuit blocks where crosstalk has to be analyzed. The transfer function $H(j\omega)$ is built with scalable R and C elements and is based on empirical equations [22] validated by a full wave analysis.

The advantage of this concept is to detect sensitivities as early as possible during the design phase in order to anticipate for potential risks. At this stage of the design, high accuracy is not necessarily required. This method then gives good indications for an optimized floor plan.

For instance, the effect of the output bit-stream of the ADC to the input of the LNA has been considered. Indeed, the voltage swing at the ADC output is close to 1.5 V peak, which might severely disturb the input signal of the LNA and degrade the sensitivity. We have shown that for a distance of 1 mm between the ADC output and the RF LNA inputs, the voltage amplitude at ADC output has been reduced down to -125 dBv at LNA input. Therefore, this effect is negligible. This method is used to optimize layout guidelines with respect to critical function performances. The isolation criterion is defined according to the maximum spurious level that can be tolerated between two circuit blocks.

The methodology has been completed by adequate measures to reduce interferences. Specific EM software [23], which can take into account heterogeneous structures, has been used to find the optimum combination of deep trenches and guard rings to improve the overall isolation between blocks. In particular, adding a deep trench to a guard ring increases the isolation by 5–8 dB at 1 GHz depending on the distance. We should remark that ac coupling through the substrate depends on its resistivity. For ac decoupling of the supply lines, we have extensively used the two top metal layers of the process with the embedded MIM capacitor to provide a good decoupling characteristic without any silicon area penalty.

Thanks to this methodology, we have reached a very compact layout without compromising the performance of the transceiver. The silicon area used for the receiver part (from the LNAs to the output bit-stream of the ADC) is only 1.8 mm^2.

IV. Experimental Results

The measured analog front-end receiver (without the ADC) NF is 2.3 dB and its IIP3 is -9 dBm, which is in agreement with the simulated results (see Section II). The dc offset is typically below 3.5 mV in all bands at the BBF output.

The measured SNR and the signal-to-noise and distortion ratio (SNDR) of a single $\Sigma\Delta$ modulator are plotted in Fig. 11. The peak SNDR is 81.8 dB and peak SNR is 82.5 dB in 135-kHz bandwidth for a single modulator. It corresponds to an effective number of bits (ENOB) of 13.5. The dynamic range (DR) is 84 dB. The IM2 and IM3 distance are 95 dB and 93 dB, respectively. Since the modulator input is limited at -2.87 dBFs, this leaves enough margin to avoid saturation. The total current consumption of ADC I and Q including the biasing is 2.8 mA under 2.5 V. The resolution and signal bandwidth corresponds to a figure of merit of $\mathrm{P}/(2^{\mathrm{ENOB}*}\mathrm{BW}) = 2.2$ pJ/conversion, which is equal to [24]. In this work, the power consumption

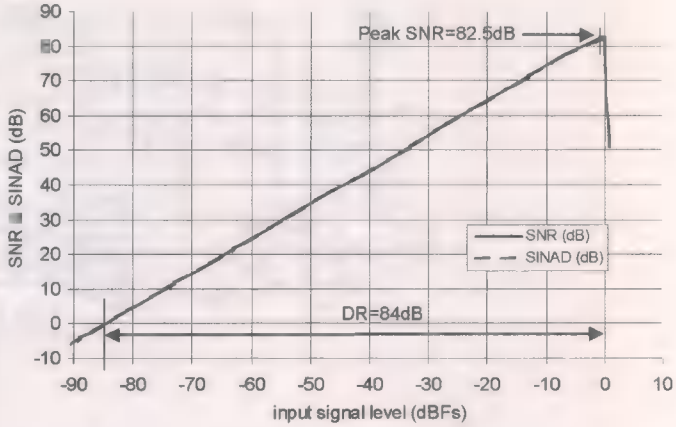


Fig. 11. Measured $\Sigma\Delta$ ADC performances as a function of input signal level.

TABLE I
Receiver Performances

Bands (MHz)	850	900	1800	1900
NF (dB)	3	3.1	3	3
LO re-radiation (dB)	-112	-118	-130	-111
AM suppression (dBm)	>-24	>-24	>-25	>-24.5
Required AM suppresion (dBm) [25]	-31	-31	-31	-31
Blocker level @ 3MHz	>-20	>-20	>-20	>-21
Required blocker level @ 3MHz *[25]*	-23	-23	-26	-26
Typical IQ phase match (°)	<1	<1	<1	<1
Typical IQ amplitude match (dB)	0.07	0.7	0.11	0.15
Typical sensitivity(dBm)	-109	-109	-108	-108
ADC dynamic range (dB)	84	84	84	84
3dB baseband BW (kHz)	208	208	208	208
Group delay to 100kHz (µs)	<0.18	<0.18	<0.18	<0.18
Rejection at 3MHz (dB)	>70	>70	>70	>70
Power (incl synth) (mA)	75	75	75	75

figure includes the I and Q $\Sigma\Delta$ modulators, the biasing circuitry, and the delay-locked loop (DLL). The DLL generates the different 13-MHz clock phase shift necessary for the RTZ clock scheme implementation. The ADC has been tested over the temperature range [$-30\,^{\circ}$C, $+85\,^{\circ}$C] and over the voltage range [2.2 V, 3.5 V] without observing any degradation in the linearity.

As illustrated in Fig. 12, the average sensitivity at 23 °C is -109, -108, and -108 dBm, respectively, for EGSM, DCS, and PCS bands. In addition, the sensitivity is not degraded at 13-MHz harmonics (dotted lines in Fig. 12), which validates the floor plan strategy detailed in Section III.

The main receiver performances are presented in Table I. The re-radiation of the LO signal measured at the LNA input is greater than -110 dBm for all bands. The I/Q quadrature, measured at the BBF output, is accurately generated with 1° and 0.2 dB. In addition, the worst case 3-MHz blocker level for a 2.4% class II RBER with a wanted signal at -98 dBm [25] can be as high as -20 dBm for GSM850/900 bands and -21 dBm for PCS1800/DCS1900 bands. This gives at least 3-dB margin for GSM850/900 and 5-dB margin for PCS1800/DCS1900 compared to the 3-MHz blocking test requirement [25]. Consequently, the LO chain exhibits good phase noise performances. The NF measured for the whole receiver at $\Sigma\Delta$ ADC output is below 3.1 dB for all bands. In the application, IP2 is verified by measuring AM suppression performance as specified in [25].

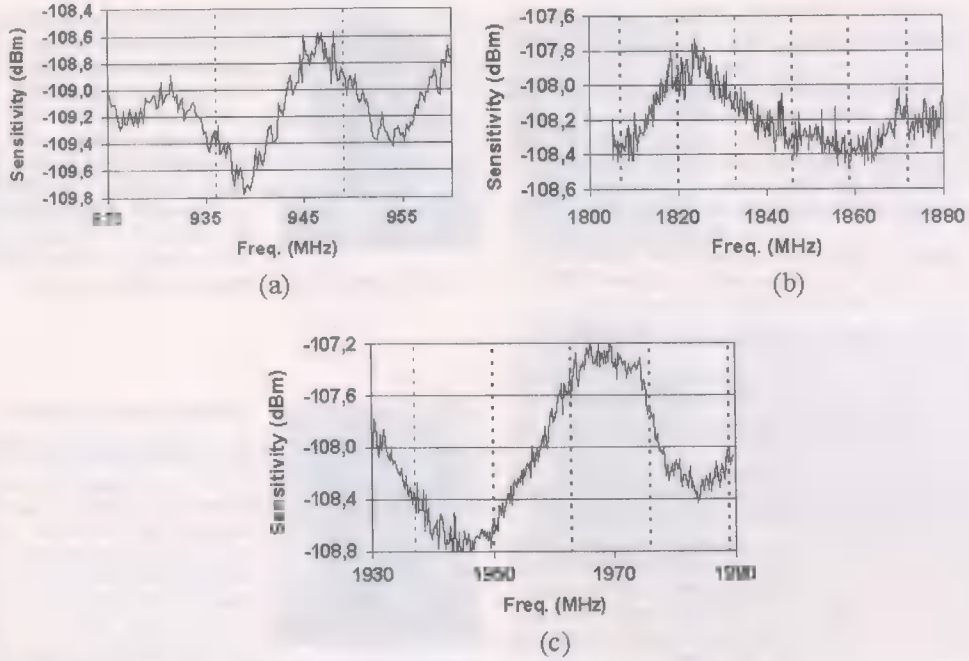


Fig. 12. Sensitivity measurement results at 23°C for EGSM (a), DCS (b) and PCS (c) bands. The dot lines represent 13 MHz harmonics.

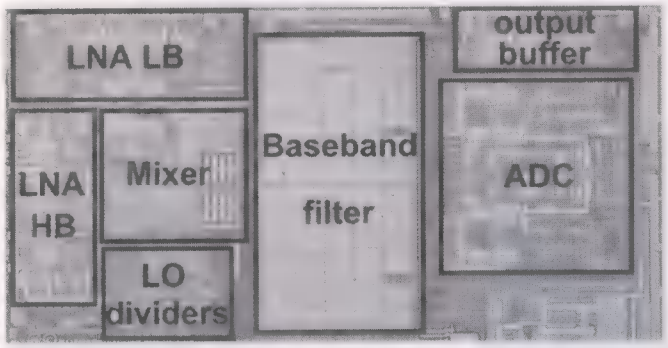


Fig. 13. Microphotograph of the receive path.

A class II RBER of 2.4% in all bands is achieved even in the presence of a −24-dBm GMSK modulated. This results in 7 dB of margin on the −31-dBm requirement [25].

The layout of the receiver is shown in Fig. 13. Its area is 1.8 mm^2.

V. Conclusion

The on-chip low-pass continuous-time ΣΔ ADC is 84-dB dynamic range over a bandwidth of ±135 kHz. Therefore, a third-order baseband filter is sufficient to attenuate the worst case blocker at 3 MHz. As a result, most of the selectivity is performed in the digital domain. In addition, the dc offset at mixer output is only amplified by 25 dB in the baseband filter and does not overload the ADC. Consequently, the dc offset cancellation is performed in the digital domain as well. The baseband buffer and ADC circuits have been enhanced to accommodate process and temperature variations. Consequently, no calibration or tuning circuitry is necessary. Moreover, a first-order substrate coupling analysis that optimizes the floor plan strategy with respect to area and crosstalk has been presented and validated since no degradation of sensitivity performance has been observed at 13-MHz harmonics. The presented quad-band GSM/GPRS/EDGE direct-conversion multimode receiver with on-chip ΣΔ ADC consumes 75 mA under 2.7 V for an area of 1.8 mm^2, making it suited to the 2.75G system solution requirements.

Acknowledgment

Special thanks are due to D. Crespo and C. Deneuchatel for layout contribution, and E. Thomas and F. Pawlus for measurement work. A special acknowledgment is extended to K. Philips, L. Breems, R. van Veldhoven, and B. Minnis from Philips Research Laboratories for fruitful technical discussions about the ADC architectures. The authors would also like to thank the anonymous reviewers for their valuable suggestions.

References

[1] A. Loke and F. Ali, "Direct conversion radio for digital mobile phones—Design issues, status, and trends," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 11, pp. 2422–2434, Nov. 2002.

[2] R. Magoo, A. Molnar, J. Zachan, G. Hatcher, and W. Rhee, "A single-chip quad-band (850/900/1800/1900 MHz) direct conversion GSM/GPRS RF transceiver with integrated VCOs and fractional-n synthesizer," *IEEE. J. Solid-State Circuits*, vol. 37, no. 12, pp. 1710–1720, Dec. 2002.

[3] B. Razavi, "Design considerations for direct conversion receivers," *IEEE. Trans. Circuits Syst. II*, vol. 44, no. 6, pp. 428–435, Jun. 1997.

[4] A. A. Abidi, "Direct-conversion radio transceivers for digital communications," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1399–1410, Dec. 1995.

[5] B. Razavi, *RF Microelectronics*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[6] B. Bastan, E. E. Bautista, G. Nagaraj, and J. Heck, "A quadrature down converter for direct conversion receivers with high 2nd and third order intercept points," in *Proc. IEEE Int. Devices, Circuits and Systems*, Mar. 2000, pp. C19/1–C19/4.

[7] F. Gatta, D. Manstetta, P. Rossi, and F. Svelto, "A fully integrated 0.18 μm CMOS direct conversion receiver front-end with on-chip LO for UMTS," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 15–23, Jan. 2004.

[8] A. A. Abidi, "General relations between IP2, IP3, and offsets in differential circuits and the effects of feedback," *IEEE Trans. Microwave Theory and Tech.*, vol. 51, no. 5, pp. 1610–1612, May 2003.
[9] D. Y. C. Lie *et al.*, "A direct-conversion W-CDMA front-end SiGe receiver chip," in *IEEE RFIC Symp.*, Jun. 2002, pp. 31–34.
[10] D. Manstretta, M. Brandolino, and F. Svelto, "Second order intermodulation mechanism in CMOS downconverters," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 394–406, Mar. 2003.
[11] H. Darabi and A. A. Abidi, "Noise in RF CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 35, no. 1, pp. 15–25, Jan. 2001.
[12] R. Brederlow, W. Weber, C. Dahl, D. Scmitt-Landsiedel, and R. Thewes, "Low frequency noise of integrated poly-silicon resistors," *IEEE J. Solid-State Circuits*, vol. 48, no. 6, pp. 1180–1187, Jun. 2001.
[13] J. A. Weldon, R. S. Narayanaswami, J. C. Rudell, L. Lin, M. Otsuka, S. Dedieu, L. Tee, K. Tsai, C. Lee, and P. R. Gray, "A 1.75-GHz highly integrated narrow-band CMOS transmitter with harmonic-rejection mixers," *IEEE J. Solid-State Circuits*, vol. 36, pp. 2003–2015, Dec. 2001.
[14] *Radio Transmission and Reception*, 3GPP TSG GSM/EDGE RAN, Release 5, 3GPP TS 45.005, V5.9.0 (2003-08).
[15] R. van Veldhoven, "A triple-mode continuous-time $\Sigma\Delta$ modulator with switched-capacitor feedback DAC for a GSM-EDGE/CDMA2K/UMTS receiver," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2003, pp. 60–61.
[16] E. J. van der Zwan, K. Philips, and C. A. A. Bastiaansen, "A 10.7 MHz if-to-baseband $\Sigma\Delta$ A/D conversion system for AM/FM radio receivers," *IEEE J. Solid-State Circuits*, vol. 35, no. 12, pp. 1810–1819, Dec. 2000.
[17] E. J. van der Zwan and E. C. Dijkmans, "A 0.2 mW CMOS $\Sigma\Delta$ modulator for speech coding with 80 dB dynamic range," *IEEE J. Solid-State Circuits*, vol. 31, no. 12, pp. 1810–1819, Dec. 1996.
[18] K. Philips, "A 4.4 mW, 76 dB complex $\Sigma\Delta$ ACD for a low-cost bluetooth receiver," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2003, pp. 64–65.
[19] Y. Le Guillou, H. Marie, P. Gamand, and P. Descamps, "Biasing technique for high dynamic range, low oversampling ratio continuous-time $\Sigma\Delta$ modulators without calibration," *IEE. Elec. Lett.*, vol. 39, no. 25, pp. 1792–1793, Dec. 2003.
[20] R. W. Adams, "Design and implementation of an adio 18-bit analog-to-digital converter using oversampling techniques," *J. Audio Eng. Soc.*, vol. 34, pp. 153–166, Mar. 1986.
[21] D. Szmyd, R. Brock, N. Bell, S. Harker, G. Patrizi, J. Fraser, and R. Dondero, "Qubic4: A silicon RF-BiCMOS technology for wireless communication Ics," in *Proc. Bipolar/BiCMOS Circuits and Technology Meeting*, Sep. 2001, pp. 60–63.
[22] D. K. Su, M. J. Loinaz, S. Masui, and B. A. Wooley, "Experimental results and modeling techniques for substrate noise in mixed-signal integrated circuits," *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 420–430, Apr. 1993.
[23] D. Bajon, S. Wane, H. Baudrand, and P. Gamand, "Full wave analysis of isolated pocket to improve isolation performances in silicon based technology," in *IEEE MTT-S Symp. Dig.*, Jun. 2002, pp. 987–990.
[24] O. Oliaei, P. Clément, and P. Gorisse, "A 5 mW sigma-delta modulator with 84-dB dynamic range for GSM/EDGE," *IEEE J. Solid-State Circuits*, vol. 37, no. 1, pp. 2–10, Jan. 2002.
[25] *Mobile Station (MS) Conformance Specification*, 3GPP TSG GSM/EDGE RAN, Release 5, 3GPP TS 51.010-1 V5.7.0 (2004-02), section 14.

Yann Le Guillou was born in France in 1974. In 1999, he received the degree of engineer from the Ecole Supérieure d'Electricité (Supélec), Paris, France. He is currently working toward the Ph.D. degree on high dynamic range modulator for A/D conversion integrated in RF transceivers.

He has been working for Philips Semiconductors, Caen, France, since 1999. He has been involved in fractional-N synthesizers and ADC design for cellular handsets.

Olivier Gaborieau was born in France in 1972. In 1996, he received the degree of engineer from the Institut de Formation d'Ingénieur en Techniques Electronique de Paris (IFITEP).

From 1993 to 1996, he was with TRT Lucent Technologies, France, where he was involved in the design of point-to-point microwave radio link up to 40 GHz. From 1996 to 1999, he worked for Wavecom, France, where he was involved in the design of IC transceivers for GSM. He joined Philips Semiconductors, Caen, France, in 2000. He is working on analog IC design for cellular products.

Patrice Gamand was born in France on March 1959. In 1984, he received the Ph.D. degree in microelectronics from the University of Lille, France.

He joined Philips Research Labs in France in 1984 to work on microwave and millimeter waves integrated ICs for radar, satellite and radio link applications. In 1993, he moved to Philips Semiconductors, Caen, France, as a design group leader to develop read/write amplifiers for HDDs. In 1998, he joined the telecommunication activity as Development Manager for Cellular RF ASICs and in 2000 he was managing the Cellular RF ASICs business. He joined the Competence Centre RF within Philips Semiconductors as Development Services Manager in 2002 and now he is the Technology Manager. He is an author or co-author of more than 20 technical papers and holds 20 patents.

Martin Isberg received the M.Sc. degree in electrical engineering from Lund University, Lund, Sweden, in 1989 and the Licentiate degree in applied electronics from Lund University in 1993.

In 1988, he joined Ericsson in Lund to work with RF development for cellular handsets. In 1993, he started to work with RF integration with focus on direct conversion and IC design. He is presently with Ericsson Mobile Platforms, Lund, where he is Development Manager for the RF and Mixed-Signal Technology Unit.

Peter Jakobsson received the M.Sc. degree in electronics from the University of Lund, Lund, Sweden, in 1995.

In 1995, he joined Ericsson as an RF Designer. Presently, he is with Ericsson Mobile Platforms, Lund, where he is an RF System Designer in the RF-IC Design Group. His field of work includes design and analysis of radio systems for both GSM and UMTS applications. He holds eight patents in the field of cellular systems.

Lars Jonsson received the M.Sc. degree in electronics from Lund University, Lund, Sweden, in 1989.

He joined Ericsson in Lund as an RF Designer for mobile telephones in 1989. During 1992–1997, he worked at the Department of Applied Electronics at Lund University with RF-IC design. He returned to Ericsson and joined the RF and Mixed-Signal Technology Unit in 1997. He is presently with Ericsson Mobile Platforms where he is an RF-IC Project Leader.

David Le Déaut was born in France in 1973. He received the M.Sc. degree in microelectronic (D.E.S.S.) from the University of Bordeaux I, Gironde, France, in 2000.

Since 2000, he has been with Philips Semiconductors, Caen, France, where he is ■ Design Engineer in the RF IC Design Group. He is currently designing direct conversion transceiver blocks for the GSM standard.

Hervé Marie received the engineer degree in electrical engineering from ENSI Caen, France, in 1988.

In 1988, he joined Philips where he designed analog-to-digital converters. During that same year, he taught IC design at ENSI Caen. From 1990 to 1994, he worked for Ion Beam Applications, Belgium, where he designed particle accelerators for medical and industrial applications. In 1994, he rejoined Philips, where he designed a variety of mixed-signals circuits such as video filters, high-speed front-ends, PLLs, and continuous-time sigma-delta modulators. He holds nine patents in circuit design. He currently leads a design group focusing on RF ICs.

Sven Mattisson received the Ph.D. degree in applied microelectronics from Lund University, Lund, Sweden, in 1986.

From 1987 to 1994, he was an Associate Professor in applied micro electronics at Lund Unviersity, where his research was focused on circuit simulation and analog ASIC design. In 1995, he joined Ericsson in Lund to work on cellular handset development. Presently, he is with Ericsson Mobile Platforms, Lund, where he is a Senior Expert in analog system design. Since 1996, he has also been an Adjunct Professor at Lund University. He is one of the principal developers of the Bluetooth concept.

Laurent Monge was born in Grenoble, France, in 1974. He received the M.Sc. degree in electrical engineering from the Institut National Polytechnique of Grenoble (INPG) in 1999.

Since 1999, he has been with Philips Semiconductors, Caen, France, developing transceiver products for the GSM and the automotive markets. He is currently involved in system definition for low-cost low-power transceivers operating in the ISM bands.

Torbjörn Olsson received the M.Sc. degree in electrical engineering from the Lund Institute of Technology, Lund, Sweden, in 1992.

In 1993, he joined Ericsson Components, Stockholm, where he designed battery-charging circuits for mobile phones. He moved back to Lund and Ericsson Mobile Communications in 1994, where he has been involved in the development of direct conversion receivers, integrated VCOs, and direct modulation transmitters. He currently leads a design group in Ericsson Mobile Platforms focusing on RFICs for GSM and 3G platforms.

Sébastien Prouet was born in France in 1976. He received the degree of engineer from the Ecole Superieure d'Electronique et de Radioelectricite de Grenoble in 2000.

In 2000, he joined Philips Semiconductor, Caen, France, as ■ Design Engineer in the RF ASICS design group. His research areas include the design and analysis of analog building blocks for RF transmitter chains intended for GSM and UMTS applications.

Tobias Tired received the M.Sc. degree in engineering physics from Lund University, Lund, Sweden, in 1992.

In 1993, he joined Ericsson Components, Stockholm, where he worked with bipolar process technology for high-voltage applications. In 1996, he joined Ericsson Mobile Communications, Lund, as an RF-IC Designer mainly for direct conversion receivers. Presently, he is with Ericsson Mobile Platforms, Lund, where he is an RF-IC Project Leader.

An Adaptive ENG Amplifier for Tripolar Cuff Electrodes

Andreas Demosthenous, *Member, IEEE,* and Iasonas F. Triantis, *Student Member, IEEE*

***Abstract*—Electroneurogram (ENG) recording from tripolar cuff electrodes is affected by interference signals, mostly generated by muscles nearby. Interference reduction may be achieved by suitably designed amplifiers such as the true-tripole and quasi-tripole systems. However, in practice their performance is severely degraded by cuff imbalance, resulting in very low output signal-to-interference ratios. Although some improvement may be offered by post filtering, this considerably increases complexity, size and power dissipation, rendering the approach unsuitable for the development of a high-performance ENG recording system which is fully implantable. This paper describes an integrated, fully implantable, adaptive ENG amplifier developed to automatically compensate for cuff imbalance, and thus significantly improve the quality of the recorded ENG. Measured results show that the adaptive ENG amplifier has a yield of 100%, a cuff imbalance correction range of more than ±40%, and an output signal-to-interference ratio of about 2/1 (6 dB) even for ±40% imbalance. The latter should be compared with an input signal-to-interference ratio of 1/500 (−54 dB). The circuit was fabricated in 0.8-μm BiCMOS technology, has a core area of 0.68 mm^2, and dissipates 7.2 mW from ±2.5 V power supplies. The adaptive ENG amplifier advances the state-of-the-art in implantable tripolar nerve cuff electrode recording techniques.**

***Index Terms*—Analog integrated circuits, cuff imbalance, ENG amplifier, implanted devices, tripolar cuff electrodes.**

I. Introduction

ELECTRONEUROGRAM (ENG) recording techniques for peripheral nerves using cuff electrodes offer a noninvasive way of obtaining information regarding nerve operation [1]. In the case of spinal cord injury this information can be used for the improvement of implanted devices used for rehabilitation. Monitoring nerve operation allows some level of intervention by means of functional electrical stimulation for partial control of organs suffering from paralysis and for blockage of unwanted sensory and/or motor signals. Applications that have been investigated include the correction of foot-drop, stimulating hand-grasp, and controlling the urinary bladder after spinal cord injury [2]–[4].

Recording ENG effectively is not a trivial task, as the microvolt-order (typically 1–5 μV) nerve signals are often obscured by the millivolt-order (typically 1 mV) electromyogram (EMG) from muscles nearby and by noise, notably white noise from the interstitial fluid and from the electrode–tissue interface [5], [6]. Furthermore, the spectra of the two signals overlap considerably, making separation by means of filtering very difficult [7]; the ENG has an energy in the 500 Hz–10 kHz band with maximum power around 1 kHz, while the EMG lies in the 1 Hz–3 kHz band and peaks at about 250 Hz [6]. Various ENG amplifier configurations make use of the properties of the cuff electrodes, mainly the *linearization* of the EMG potential field inside the cuff [8]. Improved performance in terms of EMG reduction is offered by tripolar cuffs (i.e., cuffs with three equally spaced ring electrodes embedded in the inside wall [1]). Due to this linearization, the EMG potential differences between the central electrode and the outer electrodes are equal and opposite and can be cancelled by a differential amplifier arrangement. By contrast, the ENG signal does *not* cancel in this way and can be recovered. The amplifiers used with tripolar cuffs are the *quasi-tripole* (QT) [1], [6] and *true-tripole* (TT) [9]. However, EMG reduction in these systems is affected by the departure of the cuff–tissue interface from its ideal model, caused by factors like cuff asymmetry and tissue growth inside it after implantation, resulting in *cuff imbalance* as explained in more detail in Section II.

To automatically compensate for the possible presence of cuff imbalance, and thus minimize EMG artifacts in nerve cuff electrode recording, an adaptive version of the TT, termed the *adaptive-tripole* (AT), has been proposed [10] and its first integrated realization was reported in [11]. However, the realization in [11] showed poor performance in terms of output *signal-to-interference ratio* (SIR),[1] harmonic distortion, cuff imbalance correction range, and yield. This paper describes an improved realization of the AT which overcomes all the limitations of the first design. These enhancements were necessary in order to make the system fully implantable for the targeted biomedical application (i.e., bladder implant). The adaptive ENG amplifier to be described has a chip yield of 100%, a cuff imbalance correction range of more than ±40%, and an output SIR of no less than 2 dB even for ±40% imbalance. The circuit was fabricated in 0.8-μm BiCMOS technology, occupies 0.68 mm^2, and dissipates 7.2 mW from ±2.5 V power supplies.

The remaining sections of this paper are organized as follows. In Section II, the basic principles of ENG recording from tripolar cuff electrodes are briefly reviewed. Section III describes the AT architecture and examines the effect of phase errors on system performance. Section IV describes the circuit design of the various building blocks, while measured results are presented in Section V. Finally, conclusions are drawn in Section VI.

[1]SIR refers to the ratio of the peak amplitude of the ENG signal over that of the EMG signal.

Manuscript received March 4, 2004; revised August 11, 2004. This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant GR/M88990.

The authors are with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: a.demosthenous@ee.ucl.ac.uk).

Digital Object Identifier 10.1109/JSSC.2004.840957

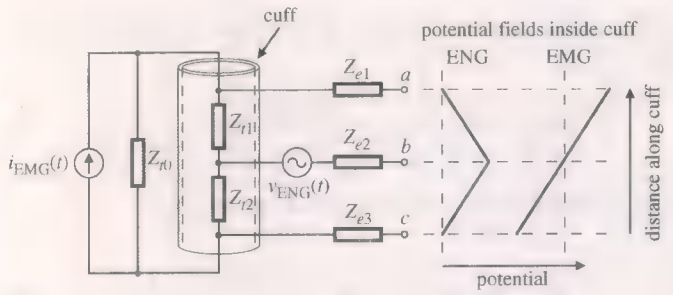


Fig. 1. Lumped-impedance model of the cuff and idealized ENG and EMG potentials inside the cuff [6], [12]. Typical impedance values: $Z_{t0} = 200\ \Omega$, $Z_{t1,2} = 1.25\ \mathrm{k}\Omega$, $Z_{e1,2,3} = 1\ \mathrm{k}\Omega$.

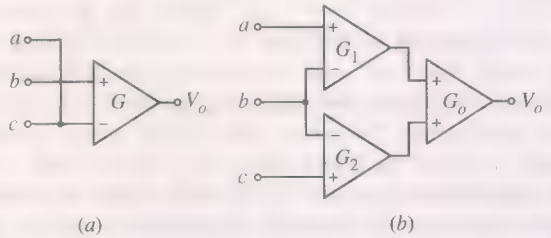


Fig. 2. Tripolar ENG amplifier configurations. (a) Quasi-tripole (QT). (b) True-tripole (TT).

II. Principles of Tripolar Cuff Electrode Recording

The ENG signal results from the action potentials propagating along the nerve fibers, which cause small action currents to flow through the fiber membranes into the extrafascicular medium [5]. Confinement within an insulating cuff causes the local impedance to be higher than outside the cuff, so that the action currents give rise to measurable potentials between the cuff electrodes. Simply stated, the nerve is an insulator, while the space between the nerve bundle and the cuff is filled with connective tissue and/or conducting fluid.

A very important function of the cuff is that, as a uniform insulating tube, any externally applied potential differences between the ends will produce a linear gradient inside [8]. This *linearization* effect is depicted in Fig. 1 in the basic electrical lumped-impedance model of the cuff [6], [12]. In this model, Z_{t1} and Z_{t2} represent the tissue impedances inside the cuff, Z_{t0} is the tissue impedance outside the cuff, Z_{e1}, Z_{e2}, and Z_{e3} are the electrode–tissue contact impedances, $i_{\mathrm{EMG}}(t)$ is the interfering EMG current that flows inside the cuff, and $v_{\mathrm{ENG}}(t)$ is the ENG voltage. At the frequencies of interest, the impedances may be regarded as purely resistive with typical values listed in the caption of Fig. 1. The EMG potentials across nodes *ab* and *cb* in Fig. 1 appear as anti-phase while the respective ENG potentials appear in-phase. Given the linear gradient of the EMG potential inside the cuff and equally spaced tripolar electrodes, the residual EMG at the output from either the QT or TT amplifier configurations (Fig. 2) will ideally be zero. However, in practice Z_{t1} and Z_{t2} are subject to uneven variations which destroy the tripolar cuff symmetry, resulting in *cuff imbalance*, defined as

$$X_{\mathrm{imb}} = \left(\frac{Z_{t1} - Z_{t2}}{Z_{t1} + Z_{t2}}\right) \times 100\%, \quad X_{\mathrm{imb}} < 100\%. \qquad (1)$$

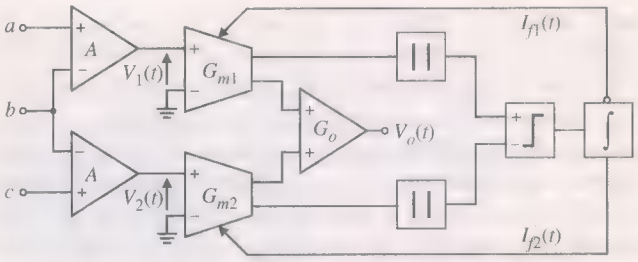


Fig. 3. Adaptive-tripole (AT) architecture.

The two main reasons for the variations in Z_{t1} and Z_{t2} are inhomogeneous tissue growth inside the cuff after implantation and manufacturing tolerances in positioning of the electrodes [10]. Secondary reasons affecting cuff imbalance include the position of the EMG source relative to the cuff [13]. Although the ENG signal recorded with the TT is about twice that recorded with the QT, the TT is much more sensitive to mismatch in Z_{t1} and Z_{t2} than the QT. On the other hand, the QT, unlike the TT, is very sensitive to mismatches in Z_{e1}, Z_{e2} and Z_{e3}. In the case of the TT, assuming unity gain for the output amplifier ($G_o = 1$), the residual EMG at its output is [14]

$$V_{o(\mathrm{EMG})} = i_{\mathrm{EMG}}(t)\left[\frac{Z_{t0}(G_1 Z_{t1} - G_2 Z_{t2})}{Z_{t0} + Z_{t1} + Z_{t2}}\right] \qquad (2)$$

where G_1 and G_2 are the gains of the input differential amplifiers in Fig. 2(b). However, note that the term on the right-hand side of (2) can be made zero by *adjusting* G_1 and G_2 to compensate for any mismatch between Z_{t1} and Z_{t2} (this approach cannot be used with the QT). An automatic adjustment of the two amplifier gains is realized by the AT, which is described in Section III.

III. Adaptive Tripole Architecture

A. System Description

The block diagram of the AT implementation described in this paper is shown in Fig. 3. The system consists of two voltage preamplifiers, each with a fixed gain A, providing a very low-noise interface with the cuff electrodes. The preamplifiers are followed by two operational transconductance amplifiers (OTAs) with variable gains G_{m1} and G_{m2}, controlled by the differential feedback currents $I_{f1}(t)$ and $I_{f2}(t)$. The control stage operates by first obtaining the *moduli* of the currents at the output of the variable-gain OTAs and applying them to a current comparator to establish which is the largest. The comparator voltage output is subsequently applied to a large time-constant integrator which generates $I_{f1}(t)$ and $I_{f2}(t)$. The variable-gain OTAs counterbalance the presence of cuff imbalance, ideally by equalizing the amplitudes of the EMG signals at their outputs. As a result, when the output signals of the OTAs are summed at the input of the output-stage amplifier (gain G_o), the equal and anti-phase EMG signals from the two channels are cancelled, and the in-phase ENG signals are added and further amplified.

B. Sensitivity to Phase Errors

The AT achieves optimum artifact reduction when the EMG terms at the inputs of the output-stage amplifier (Fig. 3) are

exactly anti-phase. However, the use of ac coupling[2] in the preamplifiers for dc offset cancellation (see Section IV-A) will, in the case of mismatched filters, introduce additional phase shifts between the composite signals $V_1(t)$ and $V_2(t)$ in Fig. 3. The phase shifts will be more pronounced on the EMG as its frequency spectrum peaks at much lower frequencies than the ENG [6]. Even if there is some phase shift between the ENG terms of $V_1(t)$ and $V_2(t)$, it will still be possible to detect neural activity in the relevant nerve bundles.

Based on the above, it is desirable to establish the maximum tolerable phase mismatch between two first-order RC high-pass filters to achieve an output SIR of no less than unity. Assuming sinusoidal signals and a phase shift $\pi + \phi$ between the EMG term in $V_2(t)$ relative to that in $V_1(t)$, then with reference to the cuff model in Fig. 1 and for $Z_{t1} > Z_{t2}$, $V_1(t)$ and $V_2(t)$ are given by

$$V_1(t) = A\left[\frac{(1+X_{\text{imb}})}{2}V_{\text{EMG}}\sin(\omega_1 t) + V_{\text{ENG}}\sin(\omega_2 t)\right] \quad (3)$$

$$V_2(t) = A\left[\frac{-(1-X_{\text{imb}})}{2}V_{\text{EMG}}\sin(\omega_1 t+\phi) + V_{\text{ENG}}\sin(\omega_2 t)\right] \quad (4)$$

where V_{ENG} and V_{EMG} are the voltage amplitudes of $v_{\text{ENG}}(t)$ and $i_{\text{EMG}}(t)[Z_{t0}(Z_{t1}+Z_{t2})/(Z_{t0}+Z_{t1}+Z_{t2})]$ in Fig. 1, respectively, and ω_1 and ω_2 their respective frequencies. Furthermore, assuming that $I_{f1}(t)$ and $I_{f2}(t)$ in Fig. 3 have settled to their final values for a given X_{imb}, such that

$$\begin{aligned} G_{m1} &= G_{mo}(1-X_{\text{imb}}) \\ G_{m2} &= G_{mo}(1+X_{\text{imb}}) \end{aligned} \quad (5)$$

where G_{mo} is the mean gain of each variable-gain OTA, the AT output is given by

$$V_o(t) = AG_{mo}G_o\left[\frac{(1-X_{\text{imb}}^2)}{2}V_{\text{EMG}}[\sin(\omega_1 t)-\sin(\omega_1 t+\phi)] + 2V_{\text{ENG}}\sin(\omega_2 t)\right] \quad (6)$$

which using standard trigonometric identities modifies to

$$V_o(t) = AG_{mo}G_o\left[\frac{(1-X_{\text{imb}}^2)}{2}V_{\text{EMG}}\sqrt{2-2\cos(\phi)} \times \cos(\omega_1 t-\theta) + 2V_{\text{ENG}}\sin(\omega_2 t)\right] \quad (7)$$

where $\theta = \tan^{-1}[(\cos(\phi)-1)/\sin(\phi)]$ is the phase shift of the residual output EMG relative to the input EMG (i.e., seen at the electrodes). Thus, if $\phi = 0$, the AT will (ideally) eliminate EMG. However, if $\phi \neq 0$, the amplitude of the residual output EMG will depend on ϕ. From (7), the output SIR can be defined as

$$\text{SIR}_{\text{out}} = \frac{4V_{\text{ENG}}}{(1-X_{\text{imb}}^2)V_{\text{EMG}}\sqrt{2-2\cos(\phi)}}. \quad (8)$$

Thus, ϕ in radians can be calculated by

$$\phi = \pm\cos^{-1}\left[1-8\left(\frac{\text{SIR}_{\text{in}}/\text{SIR}_{\text{out}}}{(1-X_{\text{imb}}^2)}\right)^2\right] \quad (9)$$

where $\text{SIR}_{\text{in}} = V_{\text{ENG}}/V_{\text{EMG}}$. For example, if $\text{SIR}_{\text{out}} = 1$, $\text{SIR}_{\text{in}} = 1/500$, and $X_{\text{imb}} = \pm 40\%$, then $\phi = \pm 0.55°$. This can be converted to an error term ε for the maximum tolerable component mismatch of two first-order RC high-pass filters. Since the ENG does not exhibit very low-frequency components, a low-end ENG amplifier bandwidth of 100 Hz is usually realized [6]. The worst case ε for a cut-off frequency of 100 Hz is when the EMG frequency is also 100 Hz, giving $\varepsilon = 1.89\%$ between the two RC product values. It should be noted that although a mismatch between the -3-dB frequency of the two filters will also introduce magnitude errors, these will be seen by the control stage of the AT as cuff imbalance and corrected.

IV. Circuit Design

A. Low-Noise Preamplifiers

The preamplifiers, being the front-end interface with the cuff electrodes, are required to exhibit very low-noise performance and have reasonable voltage gain (about 40 dB), so that low noise is not a concern for the design of the subsequent system stages. The exact gain of the preamplifiers is not important because any gain mismatch between them will be compensated for by the control stage of the AT. Thus, a simple feedforward architecture was employed as depicted in Fig. 4, thereby avoiding the complexity and noise of feedback networks. Noise optimization of the preamplifiers was explicitly described in [16], where it was shown that in order to achieve the required noise specification with minimum die area and power dissipation, the input differential pair transistors $Q1$ and $Q2$ in Fig. 4 should be bipolar. Because of this requirement, the complete adaptive ENG amplifier was implemented in BiCMOS technology, although the control stage utilizes MOS transistors only.

The preamplifier circuit in Fig. 4 consists of a simple BiCMOS OTA ($Q1$, $Q2$, $M1$, and $M2$) terminated in the load resistor R_1 (40 kΩ, V_{ref} is a dc voltage source of 0.75 V), followed by a first-order bandpass filter, which restricts the bandwidth to about 100 Hz–10 kHz. The upper cut-off frequency is obtained by the combination of resistor R_2 (500 kΩ) and capacitor C_1 (27 pF), while the lower cut-off frequency is obtained by capacitor C_2 (80 pF) with the series combination of transistors $M6$ and $M7$, the latter transistor pair forming a high value (20 MΩ) grounded linear *active* resistor. In addition to eliminating low frequencies below the ENG passband, the high-pass section of the bandpass filter also removes some of the low-frequency flicker ($1/f$) noise voltage tail and ensures a dc offset-free preamplifier output. The ac coupling mechanism

[2]In an implantable ENG amplifier, ac coupling realized by RC high-pass filters is also included in series with the cuff electrodes to prevent dc currents flowing through the tissue which would cause electrolysis, and to cancel dc offsets stemming from the electrodes [15]. However, since passive components are usually used for such filters, their cut-off frequency can be made extremely low, thereby minimizing the possibility of phase shifts.

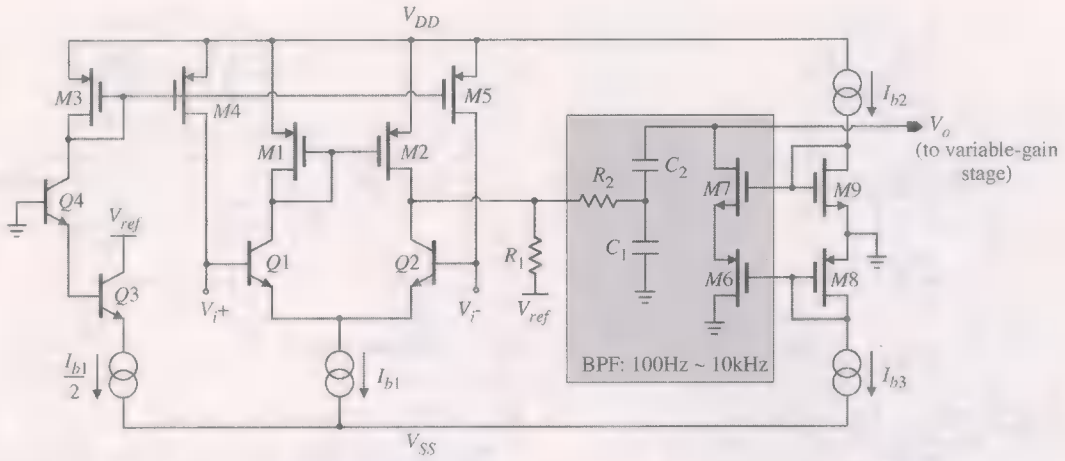


Fig. 4. Preamplifier circuit.

is very important since the succeeding variable-gain OTAs are driven single-ended, and thus, the presence of dc offset voltages (>1 mV) at their inputs would severely degrade the output SIR. By appropriate scaling of the aspect ratios of *M*6 and *M*7, a high value resistance is obtained with a maximum nonlinearity of 0.25% for a signal swing of ±85 mV. The dc bias voltages of *M*6 and *M*7 are provided by the diode-connected transistors *M*8 and *M*9, respectively, which are in turn biased by the dc current sources I_{b2} and I_{b3}.

As the base current of *Q*1 and *Q*2 cannot be supplied by the input interface, this was generated on-chip as shown in Fig. 4. Essentially, *Q*3 generates a replica of the base currents of *Q*1 and *Q*2, which is fed into the pMOS current mirror *M*3–*M*5 whose outputs feed the bases of *Q*1 and *Q*2, respectively. The base of *Q*4 is connected to ground to ensure that the emitter voltage of *Q*3 is at the appropriate level. Furthermore, the collector of *Q*3 is connected to V_{ref} to mimic as far as possible the dc conditions of *Q*1 and *Q*2 (the residual input dc base current is about 30 nA). The area of *M*4 and *M*5 were carefully chosen so that for an 800-nA drain current, their noise contribution is negligible. The bias currents for the OTA and the base current reduction circuits are provided by the dc current sources I_{b1}. The value of I_{b1} was appropriately selected so that the input-referred r.m.s. noise voltage of the preamplifier is 290 nV (noise bandwidth of 1 Hz–15 kHz). Both preamplifiers share the same current reduction and biasing circuits.

It should be noted that the preamplifiers could also be realized in CMOS technology by using the available paracitic lateral bipolar transistors. However, due to the poor matching of such devices, a larger die area and greater power dissipation would be required to meet the noise specification.

B. Variable-Gain OTAs

The composite signal at the input to each AT channel consists of EMG and ENG components with nominal peak–peak swing after preamplification of around ±50 mV (for $X_{\text{imb}} = 0$) and ±100 μV, respectively. The control stage is required to have sufficient gain to amplify the ENG to a reasonable amplitude (i.e., ±20 mV) and also sufficient linearity to accommodate the EMG

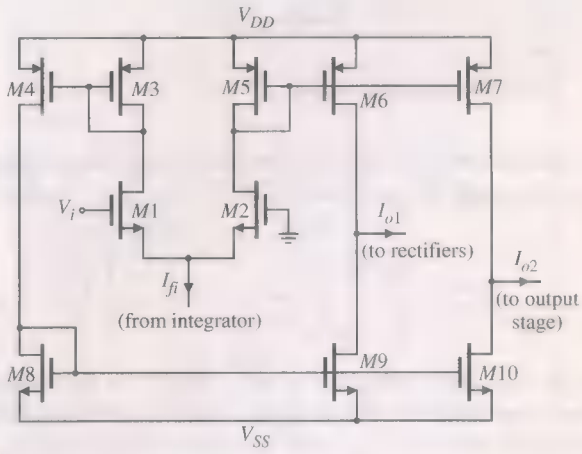


Fig. 5. Variable-gain OTA circuit.

signal. The decision to use an OTA to implement each variable-gain stage was based on the following two reasons: 1) using an OTA, variable-gain capability can be very simply achieved by changing its tail current, and 2) the output current signal from an OTA simplifies the design of the subsequent full-wave rectifiers and current comparator circuits. The basic requirement is that each variable-gain OTA must have enough linear gain range to allow even for extreme $X_{\text{imb}} = \pm 40\%$ as suggested in [13].

Although the nominal signal swing after preamplification with $X_{\text{imb}} = \pm 40\%$ is expected to be about ±70 mV, the linear input range of each variable-gain OTA was set to ±85 mV to allow for some variation in the nominal EMG amplitude picked-up from the cuff electrodes. The variable-gain OTA was designed for operation in strong inversion and its simplified schematic is shown in Fig. 5. The circuit essentially consists of a symmetrical simple CMOS OTA (input transistors *M*1 and *M*2) with current mirrors *M*3–*M*10 of unity current ratio which in practice were regulated cascodes [17]. The gain of the OTA is controlled by the feedback current I_{fi}, and the circuit has two current outputs, I_{o1} and I_{o2}, each connecting to the input of a full-wave rectifier or to the input of the output-stage amplifier.

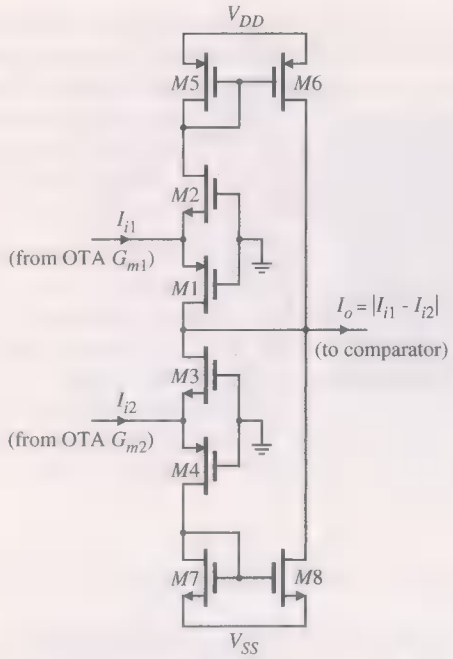


Fig. 6. Two full-wave rectifier circuits.

Assuming matched transistors and neglecting channel length modulation, each output current of the OTA in Fig. 5 is given by [18]

$$I_o = \sqrt{2I_{fi}k}V_i\sqrt{1-\frac{k}{2I_f}V_i^2}, \qquad |V_i| \le \sqrt{\frac{I_{fi}}{k}} \tag{10}$$

where V_i is the input voltage, $k = \mu C_{\rm ox}(W/2L)$ is the transconductance parameter, W and L are the channel width and length of the input transistors, μ is the carrier mobility, and $C_{\rm ox}$ is the gate oxide capacitance per unit area. The relationship between transconductance G_m and V_i can be obtained by taking the derivative of (10) with respect to V_i, yielding

$$G_m = \frac{\sqrt{2I_{fi}k}\left(1-(kV_i^2/I_{fi})\right)}{\sqrt{1-(kV_i^2/2I_{fi})}}. \tag{11}$$

For $V_i \ll \sqrt{I_{fi}/2k}$, the OTA transconductance simplifies to

$$\begin{aligned} G_m &= g_{m1,2} = \sqrt{2kI_{fi}} \\ &= \sqrt{2kI_{fo}(1 \pm 2X_{\rm imb})} \approx G_{mo}(1 \pm X_{\rm imb}) \end{aligned} \tag{12}$$

where $g_{m1,2}$ is the small-signal transconductance of transistors $M1$ and $M2$ in Fig. 5 and I_{fo} is the mean (dc) value of I_{fi}. Furthermore, in order to maintain less than 1% nonlinearity, it is required that

$$|V_i| \le 0.2\sqrt{\frac{I_{fi}}{k}}. \tag{13}$$

Given the nature of the signals after preamplification as discussed and aiming for an output-stage transimpedance gain of about 500 kΩ, a mean value for G_m of 185 μA/V was chosen. Thus, for $V_i = \pm 85$ mV, (12) and (13) can be solved for suitable values of k and I_{fi}.

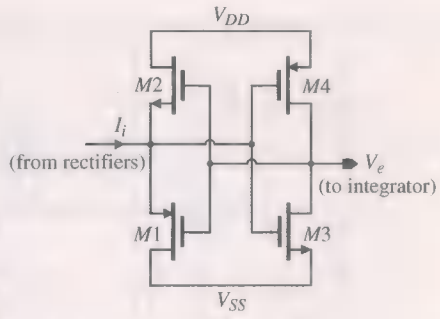


Fig. 7. Comparator circuit.

C. Full-Wave Current Rectifiers

The two full-wave rectifiers shown in Fig. 3 were realized by the current-mode circuit in Fig. 6. The upper rectifier ($M1$, $M2$, $M5$, $M6$) operates on current I_{i1} stemming from OTA G_{m1}, while the lower rectifier ($M3$, $M4$, $M7$, $M8$) operates on current I_{i2} stemming from OTA G_{m2}. The core of each current rectifier are the complementary transistors $M1$, $M2$ (upper rectifier) and $M3$, $M4$ (lower rectifier), each transistor performing half-wave precision current rectification [19]. During positive excursions of I_{i1} and I_{i2}, $M1$ and $M4$ are turned on and $M2$ and $M3$ are turned off. Thus, the drain currents of $M1$ and $M4$ equal I_{i1} and I_{i2}, respectively, while that of $M2$ and $M3$ are zero. During negative excursions of I_{i1} and I_{i2}, $M2$ and $M3$ are turned on and $M1$ and $M4$ are turned off. In this mode the drain currents of $M2$ and $M3$ equal I_{i1} and I_{i2}, respectively, while that of $M1$ and $M4$ are zero. For the upper rectifier, full-wave rectification is obtained by mirroring the drain current of $M2$ through the unity-gain pMOS current mirror $M5$, $M6$ and adding the mirror output to the drain current of $M1$. Similarly for the lower rectifier, full-wave rectification is obtained by mirroring the drain current of $M4$ through the unity-gain nMOS current mirror $M7$, $M8$ and adding the mirror output to the drain current of $M3$. In practice both current mirrors were realized by regulated cascodes [17]. The addition of the various drain currents is done at the input node of the current comparator, resulting in the output current $I_o = |I_{i1} - I_{i2}|$ as indicated in Fig. 6. Although a considerable voltage drop of about 2 V is generated at the input node of each rectifier, the use of regulated cascode mirrors with long transistors in the variable-gain OTAs, ensures that I_{i1} and I_{i2} are not degraded by channel length modulation.

D. Current Comparator

The output currents from the two full-wave rectifiers are summed at the input of the current comparator circuit [20] shown in Fig. 7 to form current I_i. The comparator uses a CMOS inverter ($M3$, $M4$) to apply negative feedback around a class-B voltage buffer ($M1$, $M2$). As a result of the feedback, the comparator input has a low-impedance (in general) and is thus ideal for determining the polarity of I_i. On the other hand, the output of the inverter does not swing between the power supplies and so some static power dissipation is present. Fortunately, since in this application low-speed operation is required, the inverter transistors can be scaled to minimize power dissipation. The buffer transistors have zero dc power dissipation.

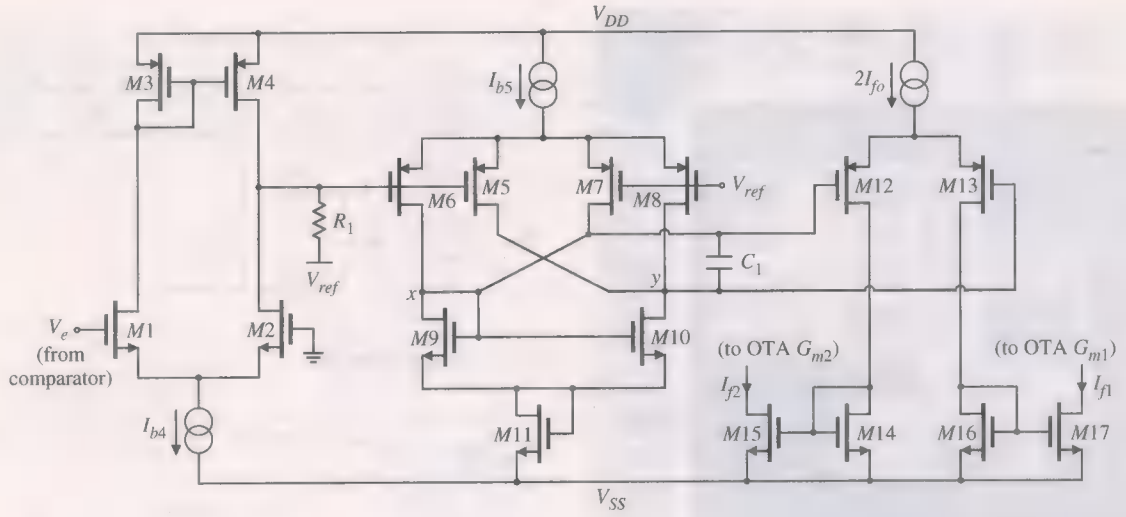


Fig. 8. Large time-constant integrator circuit.

E. Large Time-Constant Integrator

Because of the nature of cuff imbalance variations as discussed in Section II, the integrator time-constant should be as large as possible. System level simulations have shown that a time-constant of about 1 s is required for this application. The integrator schematic is shown in Fig. 8. The circuit comprises three stages. The first stage, consisting of the simple CMOS OTA (*M*1–*M*4) terminated in resistor R_1 (2 kΩ), is essentially an attenuator which also corrects amplitude variations between the comparator peak-positive and peak-negative output voltages. This is very important since significant comparator output offsets would affect the settling time and $\mathrm{SIR_{out}}$ of the AT differently for positive and negative X_{imb} values. The second stage is the actual OTA-C integrator (operated in weak inversion), and this consists of a CMOS OTA (*M*5–*M*11) utilizing transconductance cancellation [21], and an integrating capacitor C_1 (47.5 pF) which is connected across the low and high impedance nodes x and y, respectively. The attenuation provided by the first stage ensures that the input voltage to the second-stage OTA is within its linear range of operation. The second-stage OTA is biased to achieve a transconductance G_{mc} of 6.9 nA/V given by $g_{m6,8} \times [(n-1)/(n+1)]$, where $g_{m6,8}$ is the small-signal transconductance of *M*6 and *M*8, and n is the ratio of the transconductance of *M*6 to *M*5 (or *M*8 to *M*7). Transistor *M*11 performs level-shifting of the output voltage for interfacing with the third stage.

The third stage (*M*12 – *M*17) is another transconductance stage converting the voltages across C_1 to the differential feedback currents I_{f1} and I_{f2}. The tail currents of the three integrator stages are provided by the dc current sources I_{b4}, I_{b5} and $2I_{fo}$. The OTA-C stage, being *lossy*, has the following transfer function:

$$T(s) = \frac{G_{mc}}{s2C_1 + g_o} \tag{15}$$

where s is the Laplace operator, and g_o is the small-signal output conductance seen into node y. The integrator time-constant is $2C_1/g_o$ and g_o is set by I_{b5}. Any possible dc offset voltages across nodes x and y resulting from transistor mismatches may be externally corrected by adjusting the dc voltage level of R_1 (e.g., by means of current injection). The dc voltage source V_{ref} is the same as in Fig. 4 and the simulated dc gain G_{mc}/g_o of the OTA-C stage is about 73. The integrator described here offers a simpler implementation than the design in [22], both addressing the same application.

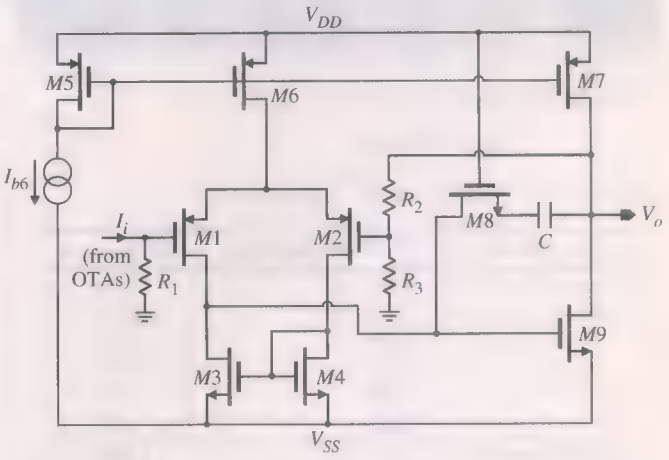


Fig. 9. Output-stage amplifier circuit.

F. Output-Stage Amplifier

The schematic of the output stage amplifier is shown in Fig. 9. The second output branch from each variable-gain OTA (Fig. 5) is hardwired to resistor R_1 (50 kΩ) where the two composite currents I_{i1} and I_{i2} are summed up to form I_i. Due to the corrective action of the control stage, the two EMG components in I_{i1} and I_{i2} are ideally of the same amplitude, and, being anti-phase, when added are cancelled out. On the other hand, the ENG components in I_{i1} and I_{i2} being in-phase, when added a voltage is generated across R_1 which is further amplified. The amplifier (*M*1–*M*9) in Fig. 9 is a standard two-stage op-amp configured as a noninverting amplifier through the feedback resistive network R_2 (90 kΩ) and R_3 (10 kΩ). The amplifier employs zero-pole compensation realized by the series combination of transistor *M*8 and capacitor C_1 (3.5 pF), and the circuit is biased

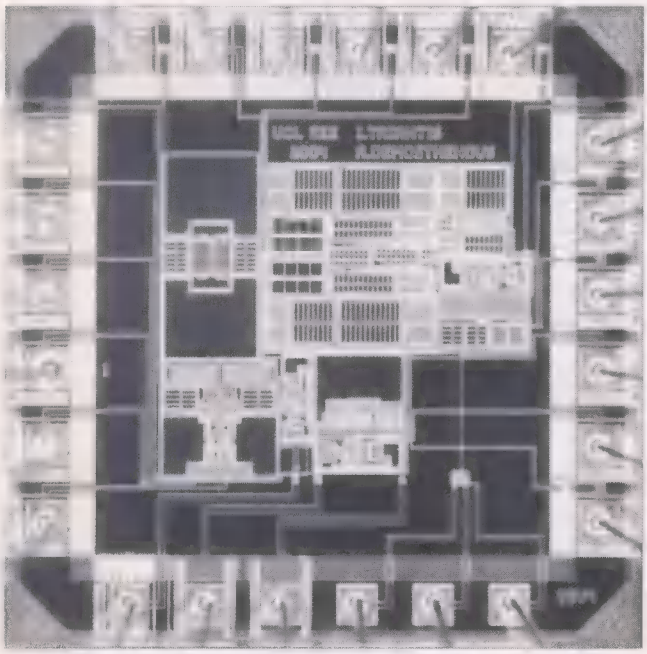


Fig. 10. Chip microphotograph.

TABLE I
MOS TRANSISTOR DIMENSIONS

Circuit	Transistor Label	W/L (μm/μm)
Preamplifier (Fig. 4)	M1, M2	150/10
	M3 – M5	20/10
	M6	2/519
	M7	2/371
	M8	7/7
	M9	5/8
Variable-gain OTA (Fig. 5)	M1, M2	150/30
Rectifiers (Fig. 6)	M1, M4	200/0.8
	M2, M3	80/0.8
Comparator (Fig. 7)	M1, M4	5/2
	M2, M3	2/2
Integrator (Fig. 8)	M5, M7	10/10
	M6, M8	10/9
	M9, M10	25/50
	M11	2/30
	M12, M13	200/5

by the dc current source I_{b6}. The simulated open-loop gain of the op-amp is 106 dB, and the input-referred r.m.s. noise current of the complete trasimpedance stage is about 100 pA (bandwidth of 1 Hz–15 kHz).

V. MEASURED RESULTS

The adaptive ENG amplifier chip, shown in Fig. 10, was fabricated in the austriamicrosystems 0.8-μm BiCMOS process [23] which includes a high resistive layer. A second chip containing the control stage configured as test structures was also fabricated. The substrates of all transistors were connected to their respective power supply rail (i.e., nMOS to V_{SS} and pMOS to V_{DD}), and the dc bias current sources I_{b1} (150 μA), I_{b2} (10 μA), I_{b3} (10 μA), I_{b4} (2 μA), I_{b5} (10 nA), $2I_{fo}$ (200 μA), and I_{b6} (50 μA), in Figs. 4, 8, and 9, were realized by an on-chip biasing circuitry (not described). Some of the key MOS transistor dimensions are listed in Table I. In total, 40 chips were fabricated (20 test structures and 20 complete systems); all showed correct operation.

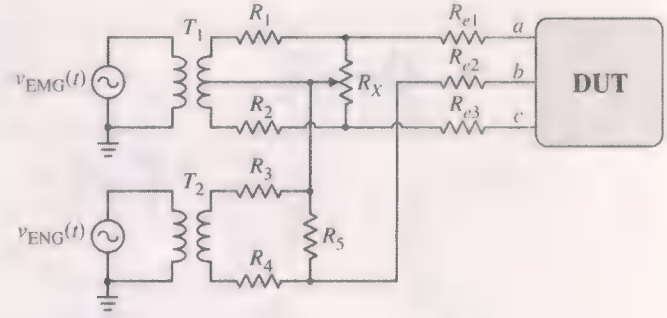


Fig. 11. Experimental setup.

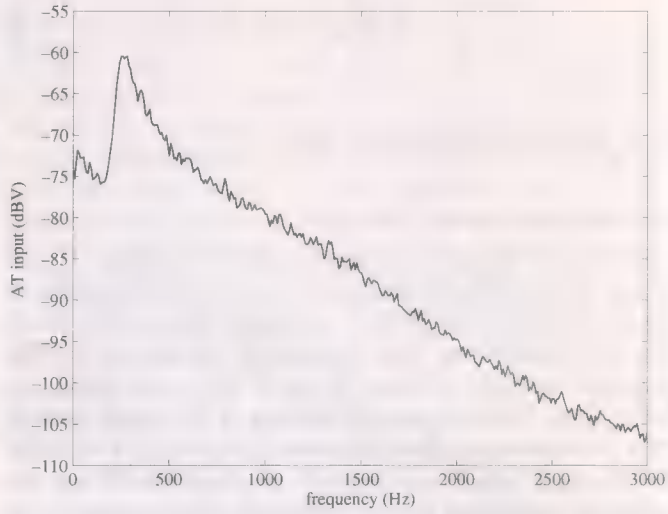


Fig. 12. Frequency spectrum of the composite input signal. The spectrum of the band-limited white noise signal representing the EMG resembles that of the real EMG.

The input ac signals to the AT chip (DUT) were provided by two audio transformers T_1 and T_2 (A262A7E) as illustrated in Fig. 11. The ac voltage sources, $v_{\rm EMG}(t)$ and $v_{\rm ENG}(t)$, generate the EMG and ENG signals, respectively, resistors R_1, R_2, R_3, R_4, R_5, and R_X provide attenuation, and the variable resistor R_X also generates amplitude imbalance (modeling $X_{\rm imb}$) between the EMG terms of the two composite signals across nodes *ab* and *cb*. Furthermore, resistors $R_{e1,2,3}$ represent the electrode resistances. Initially, the chips were tested with sinusoidal signals, $v_{\rm EMG}(t)$ (100 Hz) and $v_{\rm ENG}(t)$ (1 kHz), with nominal peak amplitudes across nodes *ab* (and *bc*) in Fig. 11 of $V_{\rm EMG} = 0.5$ mV and $V_{\rm EMG} = 1$ μV, respectively. Subsequently, in order to model a more realistic test, $v_{\rm EMG}(t)$ was replaced by an arbitrary signal (generated from band-limited Gaussian noise) with the frequency spectrum plotted in Fig. 12 (measured across *ac*). The frequency content of this signal varies between 1 Hz and 3 kHz, with a peak at approximately 250 Hz, which is the case with the real EMG signal [6]. The $v_{\rm ENG}(t)$ was kept in all measurements as a sinusoid with the characteristics mentioned above. In Fig. 12, the ENG magnitude (-114 dB) is buried under the spectrum floor of the random EMG signal.

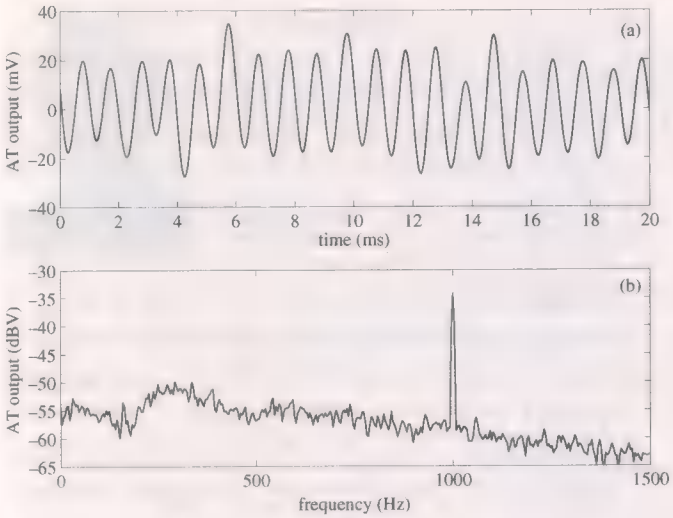


Fig. 13. System output for +20% imbalance. (a) Time-domain. (b) Frequency-domain.

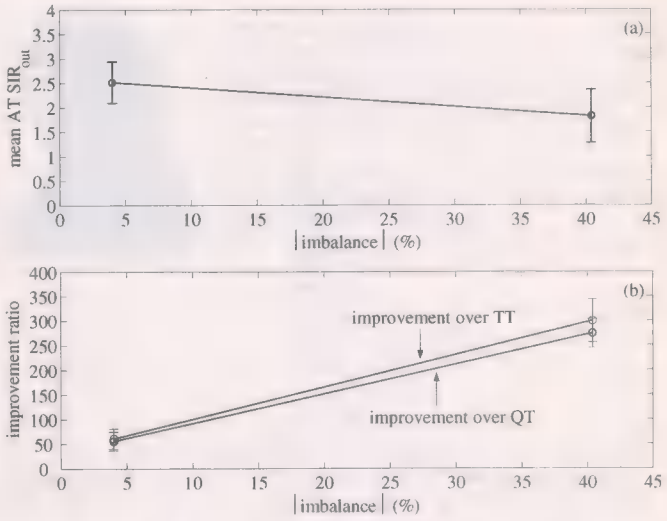


Fig. 15. (a) Mean AT $\mathrm{SIR}_{\mathrm{out}}$ versus (absolute) imbalance for all 20 chips. (b) $\mathrm{SIR}_{\mathrm{out}}$ improvement over the ideal TT and QT counterparts versus (absolute) imbalance.

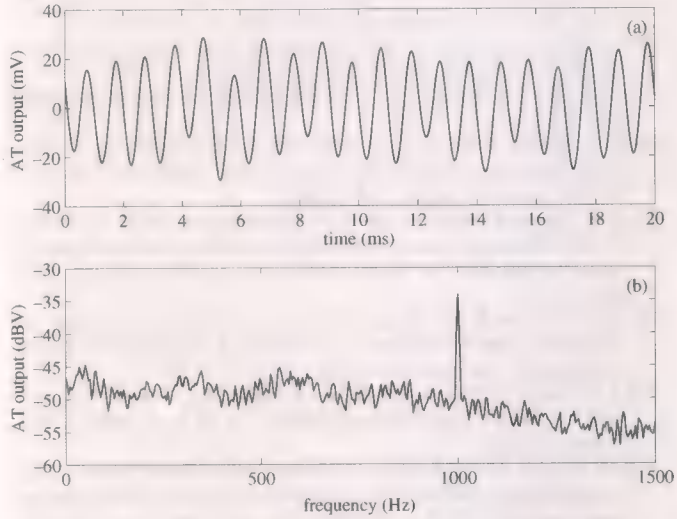


Fig. 14. System output for −40% imbalance. (a) Time-domain. (b) Frequency-domain.

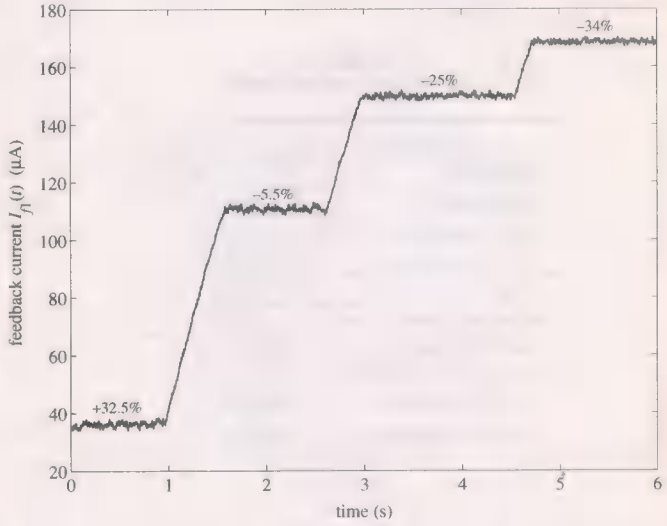


Fig. 16. Settling time of feedback current $I_{f1}(t)$ for abrupt changes in imbalance.

The time-domain tests were monitored on an Agilent 54835A Infiniium™ oscilloscope, and the frequency-domain tests on a Stanford Research Systems SR760 FFT spectrum analyzer.

Figs. 13 and 14 show the time-domain and frequency-domain outputs of the AT (after settling) for +20% and −40% imbalance, respectively. The spectra show that the $\mathrm{SIR}_{\mathrm{out}}$ is better than 3 (9.54 dB) even for 40% imbalance. This should be compared with a $\mathrm{SIR}_{\mathrm{in}}$ of 1/500 (−54 dB). These results show the superiority of the AT relative to any filtering technique because its operation is not frequency related. The average $\mathrm{SIR}_{\mathrm{out}}$ for all 20 (complete) AT chips as a function of imbalance is plotted in Fig. 15(a) (Matlab best linear-fit), where it can be seen that even for extreme values of imbalance, the mean AT $\mathrm{SIR}_{\mathrm{out}}$ is better than 2 (6 dB). The error bars in the plot indicate the spread of values from all 20 chips. For comparison, Fig. 15(b) shows the mean $\mathrm{SIR}_{\mathrm{out}}$ improvement over the theoretical TT and QT amplifier configurations as a function of imbalance (for the TT the input amplifiers were assumed to be matched, and for the QT the electrode impedance values listed in the caption of Fig. 1 were assumed). From the plot, it is apparent that the AT significantly outperforms both counterparts in the presence of imbalance. Fig. 16 shows the settling time of the feedback current $I_{f1}(t)$ in Fig. 3 for abrupt step-like changes in imbalance. The imbalance was changed successively between +32.5%, −5.5%, −25%, and −34%. The corresponding settling time (to 1%) is about 20 ms per percent change in X_{imb}.

Finally, in order to test the sensitivity of the AT architecture to phase variations, phase shifts were introduced between the two input EMG terms to the system (the additional test structure chip was used for this test). Fig. 17 shows the $\mathrm{SIR}_{\mathrm{out}}$ as a function of phase shift for both measured and theoretical cases, the latter calculated from (9) and for 40% imbalance. The two graphs show excellent agreement, but for phase values near the origin, the theoretical $\mathrm{SIR}_{\mathrm{out}}$ tends to infinity, which would never be the case for a practical realization. Saline-bath testing of the AT chip (not described here) also confirmed its high performance. The main design features of the AT chip are summarized in Table II.

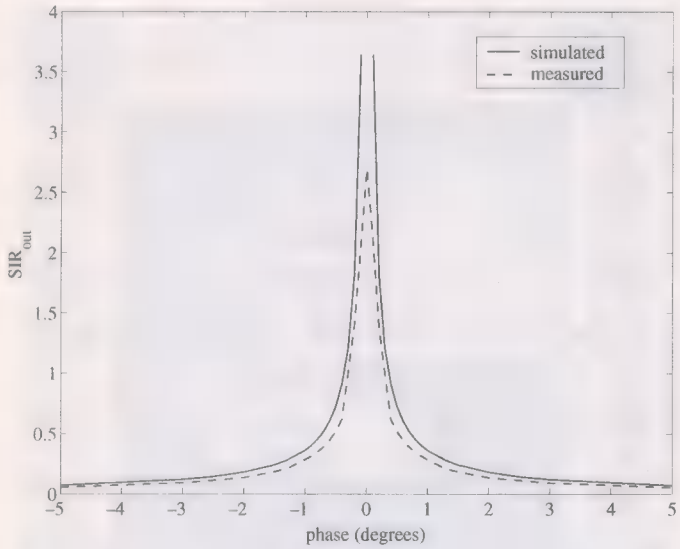


Fig. 17. Sensitivity of AT SIR_{out} to phase shifts.

TABLE II
SUMMARY OF PERFORMANCE

Parameter	Value
Technology	0.8 μm BiCMOS
Power supply	±2.5 V
Power consumption	7.2 mW
Active area (core)	0.68 mm^2
SIR_{out}	> 6dB
Imbalance correction range	> ±40%
Total ENG path gain	87dB
Setting time (step-change)	
±20 % imbalance	480 ms
±40 % imbalance	960 ms

VI. CONCLUSION

The design of an adaptive ENG amplifier for interface to tripolar cuff electrodes has been described. The adaptive ENG amplifier offers a fully implantable solution to the problem of cuff imbalance, thereby significantly advancing the state-of-the-art in the field. The described realization overcomes many of the limitations of a previous design in terms of reliability, cuff imbalance correction range, output SIR and output signal distortion. The operation of the circuit has been thoroughly verified by tests on 40 fabricated chip samples, all exhibiting correct behavior. Although the described adaptive ENG amplifier has been developed for a next-generation bladder implant, it can also be seen as a generic high-performance ENG amplifier for any functional electrical stimulation application employing tripolar nerve cuff electrodes.

ACKNOWLEDGMENT

The authors would like to thank Mr. P. Langlois for his help with the high-resistance circuit and Prof. N. Donaldson for useful discussions on the medical aspects of this work.

REFERENCES

[1] J. J. Struijk, M. Thomsen, J. O. Larsen, and T. Sinkjaer, "Cuff electrodes for long-term recording of natural sensory information," *IEEE Eng. Med. Biol. Mag.*, vol. 18, no. 3, pp. 91–98, May/Jun. 1999.

[2] T. Sinkjaer, M. K. Haugland, and J. Haase, "Natural neural sensing and artificial muscle control in man," *Exp. Brain Res.*, vol. 98, pp. 542–545, 1994.

[3] M. Haugland, A. Lickel, J. Haase, and T. Sinkjaer, "Control of FES thump force using slip information obtained from the cutaneous electroneurogram in quadriplegic man," *IEEE Trans. Rehab. Eng.*, vol. 7, no. 2, pp. 215–227, Jun. 1999.

[4] M. Hansen, M. Haugland, T. Sinkjaer, and N. Donaldson, "Real time foot drop correction using machine learning and natural sensors," *Neuromodulation*, vol. 5, pp. 41–43, Jan. 2002.

[5] R. Stein, C. Davis, A. Jhamanda, T. Mannard, and T. Nichols, "Principles underlying new methods for chronic neural recording," *Le J. Canadien des Sciences Neurologiques*, vol. 2, pp. 235–244, 1975.

[6] Z. M. Nikolić, D. B. Popović, R. B. Stein, and Z. Kenwell, "Instrumentation for ENG and EMG recordings in FES systems," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 7, pp. 703–706, Jul. 1994.

[7] B. Upshaw and T. Sinkjaer, "Digital signal processing algorithms for the detection of afferent nerve activity recorded from cuff electrodes," *IEEE Trans. Rehab. Eng.*, vol. 6, no. 2, pp. 172–181, Jun. 1998.

[8] J. Struijk and M. Thomsen, "Tripolar nerve cuff recording: Stimulus artifact, EMG, and the recorded nerve signal," in *Proc. 17th Int. Conf. IEEE EMBS*, vol. 2, Montreal, Canada, 1995, pp. 1105–1106.

[9] C. Pflaum, R. R. Riso, and G. Wiesspeiner, "Performance of alternative amplifier configurations for tripolar nerve cuff recorded ENG," in *Proc. 18th Int. Conf. IEEE EMBS*, vol. 1, Amsterdam, The Netherlands, 1996, pp. 375–376.

[10] M. S. Rahal, "Optimization of nerve cuff electrode recordings for functional electrical stimulation applications," Ph.D. dissertation, University College London, London, U.K., 2000.

[11] I. F. Triantis, R. Rieger, J. Taylor, A. Demosthenous, and N. Donaldson, "A CMOS adaptive interference reduction system for nerve cuff recordings," in *Proc. 28th Eur. Solid-State Circuits Conf.*, Florence, Italy, 2002, pp. 113–116.

[12] M. Sahin and M. D. Durand, "An interface for nerve recording and stimulation with cuff electrodes," in *Proc. 19th Int. Conf. IEEE EMBS*, Chicago, IL, 1997, pp. 2004–2005.

[13] I. F. Triantis, A. Demosthenous, and N. Donaldson, "On cuff imbalance and tripolar ENG amplifier configurations," *IEEE Trans. Biomed. Eng.*, to be published.

[14] M. Rahal, J. Winter, J. Taylor, and N. Donaldson, "An improved configuration for the reduction of EMG in electrode cuff recordings: A theoretical approach," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1281–1284, Sep. 2000.

[15] K. Papathanasiou and T. L. Ehmann, "An implanatable CMOS signal conditioning system for recording nerve signals with cuff electrodes," in *Proc. ISCAS 2000*, vol. 5, Geneva, Switzerland, pp. 281–284.

[16] R. Rieger, J. Taylor, A. Demosthenous, N. Donaldson, and P. Langlois, "Design of a low-noise preamplifier for nerve cuff electrode recording," *IEEE J. Solid State Circuits*, vol. 38, no. 8, pp. 1373–1379, Aug. 2003.

[17] E. Sackinger and W. Guggenbuhl, "A high-swing, high-impedance MOS cascode circuit," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 289–298, Feb. 1990.

[18] A. Nedungadi and T. R. Viswanathan, "Design of linear CMOS transconductance elements," *IEEE Trans. Circuits Syst.*, vol. CAS-31, no. 10, pp. 891–894, Oct. 1984.

[19] Z. Wang, "Novel pseudo RMS current converter for sinusoidal signals using a CMOS precision current rectifier," *IEEE Trans. Instrum. Meas.*, vol. 39, no. 4, pp. 670–671, Aug. 1990.

[20] H. Träff, "Novel approach to high-speed CMOS current comparators," *Electron. Lett.*, vol. 28, pp. 310–312, Jan. 1992.

[21] P. Garde, "Transconductance cancellation for operational amplifiers," *IEEE J. Solid-State Circuits*, vol. 12, no. 6, pp. 310–311, Jun. 1977.

[22] R. Rieger, A. Demosthenous, and J. Taylor, "A 230-nW 10-s time constant integrator for an adaptive nerve signal amplifier," *IEEE J. Solid State Circuits*, vol. 39, no. 11, pp. 1968–1975, Nov. 2004.

[23] "0.8 μm BiCMOS Process Parameters," AustriaMicroSystems (AMS) Int. AG, Doc. 9933008, 2.0 ed., 2001.

Andreas Demosthenous (S'96–M'99) was born in Nicosia, Cyprus, in 1969. He received the B.Eng. degree in electrical and electronic engineering from the University of Leicester, Leicester, U.K., in 1992, the M.Sc. degree in telecommunications technology from Aston University, Birmingham, U.K., in 1994, and the Ph.D. degree in electronic and electrical engineering from University College London (UCL), London, U.K., in 1998.

From 1998 to 2000, he held a Postdoctoral Research Fellow position in the Department of Electronic and Electrical Engineering, UCL. In September 2000, he was appointed as a Lecturer in the same department, where he heads the Analog Electronics research group. His main area of research is analog and mixed-signal integrated circuits for biomedical, digital communications, and video signal processing applications.

Iasonas F. Triantis (S'03) was born in Geneva, Switzerland, in 1976. He received the M.Eng. degree from the Department of Electrical Engineering and Electronics, University of Manchester Institute of Science and Technology, U.K., in 2000. From 2000 to 2003, he was a Research Assistant in the Department of Electronic and Electrical Engineering, University College London, where he is currently pursuing the Ph.D. degree.

His main interests include analog IC design and medical electronics for implanted devices.

Noise-Shaping Techniques Applied to Switched-Capacitor Voltage Regulators

Arun Rao, William McIntyre, *Member, IEEE*, Un-Ku Moon, *Senior Member, IEEE*, and Gábor C. Temes, *Life Fellow, IEEE*

***Abstract*—A delta-sigma control loop for a buck-boost dc-dc converter with fractional gains is presented. This technique reduces the tones caused by the traditional pulse-frequency modulation regulation. The prototype regulator was fabricated in a 0.72-μm CMOS process and clocked at 1 MHz. It achieved suppression of tones up to 55 dB in the 0–500-kHz range. The input voltage range was 3–5 V. The output voltage ranged from 1.8 to 4 V for load currents up to 150 mA.**

***Index Terms*—Boost, buck, dc–dc converter, delta-sigma, noise shaping, voltage regulators.**

I. Introduction

SMALL electronic devices are commonly powered by batteries, which allow them to be portable. However, as battery use continues, the battery voltage drops, sometimes gradually and sometimes suddenly, depending on the type of battery and type of electronic device. Such variations in the battery voltage may have undesirable effects on the operation of the device powered by the battery. Also, the battery voltage may not be optimal for the device. Consequently, dc–dc converters are used to provide a stable output supply voltage of suitable magnitude from the battery to the electronic device.

For many years, the inductive conversion topology has been the standard way to provide a stable voltage from a battery. With the continued shrinking of handheld devices such as cell phones, PDAs, pagers and laptops, the use of inductive regulators is becoming less attractive. A compact switched-capacitor (SC) regulator is preferable to the bulky inductive regulator. SC power conversion offers reduced physical volume, less radiated EMI, as well as efficiency and cost advantages over inductive based structures. A fixed gain SC dc–dc *boost converter* may have a gain greater than or equal to one, while a fixed gain SC dc–dc *buck converter* may have a gain less than or equal to one.

In addition to increasing or decreasing the battery voltage, voltage regulation is required to maintain the battery voltage at a constant desired value. A conventional method to regulate voltage in a SC converter is to use pulse-frequency modulation (PFM) or burst-mode operation. These control techniques suffer from tones in the frequency spectrum. The tones are difficult to filter out, as their frequencies vary with load and input voltage. As a result, circuits that use the regulated voltage are susceptible to tones in the frequency region of operation. Furthermore, these tones can mix with unwanted signals outside the band of interest and modulate into the desired signal band.

In this paper, an alternate control technique using a delta-sigma loop is presented [1], which spreads the tones of the conventional SC regulator. The charge pump used to convert the input voltage acts as a D/A converter in the loop, and its output ripple is frequency shaped by the delta-sigma control loop, which also provides the pulse-frequency modulation needed for the conversion. We have applied the new control loop architecture successfully to an existing buck-boost fractional-gain regulator [2]. We could potentially inject a long pseudo-random sequence into the existing PFM loop but we then have no control over the PFM part of it. We cannot randomly make the regulator "skip" or "pump" based on a pseudo-random sequence. We would need some information of the output and input (for gain selection between the 7 different switch capacitor gains), and that will then introduce tones as it will be similar to the PFM type architecture. Using the delta-sigma control makes it possible to incorporate the gain selection into the control loop, thus providing noise shaping along with PFM control in a very small area. The measured results indicated that the tones generated by the burst-mode regulation circuitry can be reduced by as much as 55 dB by embedding the dc–dc converter in a delta-sigma loop. This verified the usefulness of the proposed scheme. It should be noted that the tones are reduced by 55 dB with respect to the noise floor of the PFM pump. The noise floor of the regulator with the delta-sigma control will be higher, because the total noise power remains the same as we do not filter the noise shaped spectrum (as done in a conventional delta-sigma modulator). The idea however is to convert the tones to white noise and prevent them from modulating into the audio band. The experimental results confirm the validity of the method [1].

II. Fractional Gain Setting Charge Pump Architecture

The block diagram of a widely used burst-mode switched-capacitor dc–dc voltage regulator [2] is shown in Fig. 1. The circuit contains two feedback loops. One of them is the PFM loop which compares the output voltage V_{out} with the desired output value V_{desired}, and turns the gated clock signal *on* or *off* depending on the result of the comparison. The other loop performs gain hopping. It sets the gain G to a value that it is sufficiently large to prevent reverse current flow into the battery, but

Manuscript received August 21, 2003; revised April 6, 2004. This work was supported by the National Semiconductor Corporation.

A. Rao was with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331-3211 USA. He is now with National Semiconductor, Grass Valley, CA 95945 USA.

W. McIntyre is with National Semiconductor, Grass Valley, CA 95945 USA.

U. Moon and G. C. Temes are with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331-3211 USA (e-mail: moon@eecs.oregonstate.edu).

Digital Object Identifier 10.1109/JSSC.2004.840986

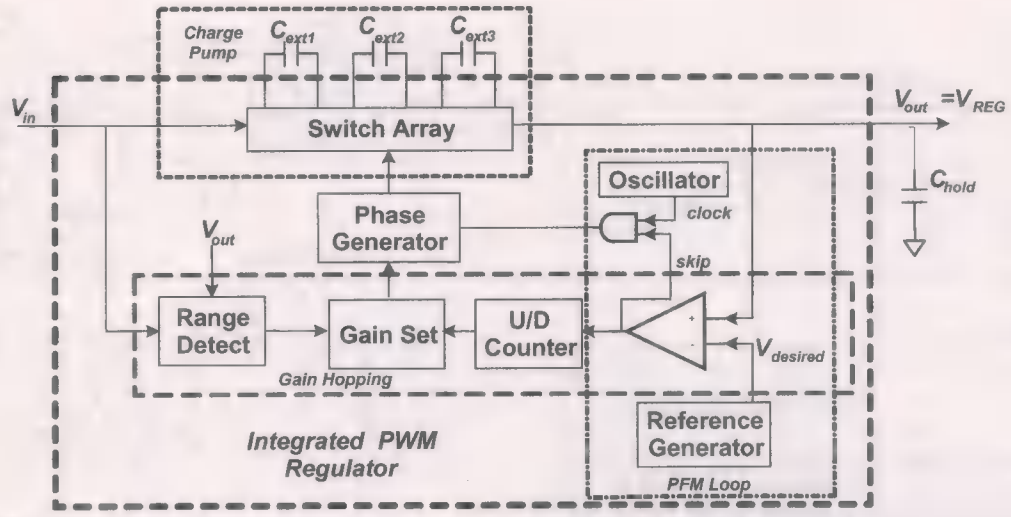


Fig. 1. Burst-mode switched-capacitor dc–dc regulator.

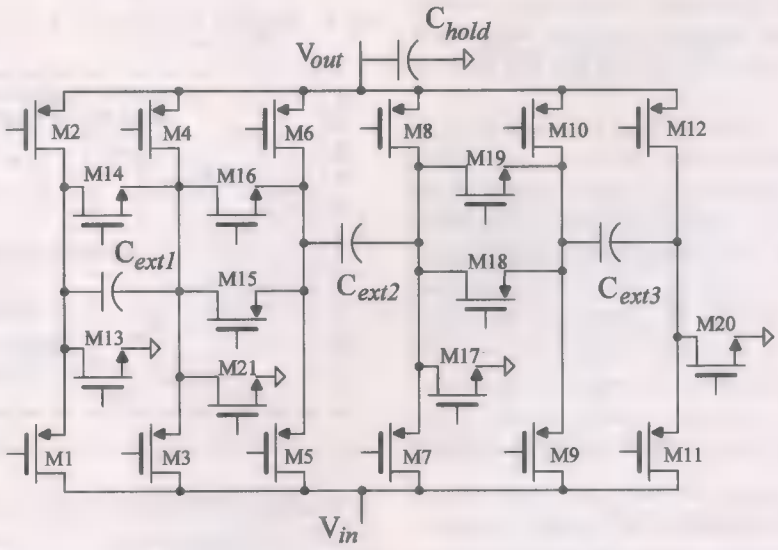


Fig. 2. Switch array with external capacitors.

not too large because then the regulator must drop the voltage by a large amount, reducing the power efficiency. The gain hopping loop requires a fractional gain setting circuit, to be discussed next.

Fractional gains can be realized by connecting external capacitors to an on-chip switch array, as shown in Fig. 2 [2]. The switch array can provide seven different gains $G = 1/2, 2/3, 3/4, 1, 4/3, 3/2$, and 2. Each gain is implemented in the two phases of a 1-MHz clock. For example, Fig. 3 shows the configuration used to implement $G = 3/2$.

To guarantee that current does not flow into the battery, we have to ensure that $G > V_{\mathrm{REG}}/V_{\mathrm{IN}}$, where V_{REG} is the desired output voltage, and V_{IN} is the unregulated battery voltage. Also, to maximize efficiency, G must be as close to $V_{\mathrm{REG}}/V_{\mathrm{IN}}$ as possible. The gain that satisfies these conditions is defined as the minimum gain G_{MIN}.

When the pump provides the gain G_{MIN}, the largest current that it can deliver to the load is approximately

$$I_{\max} \cong (G_{\min} V_{\mathrm{in}} - V_{\mathrm{reg}})/R_{\mathrm{out}} \quad (1)$$

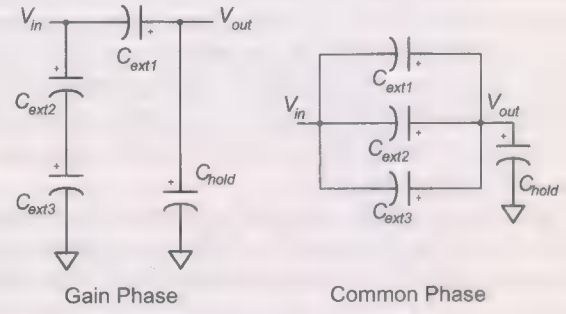


Fig. 3. Capacitor configuration for gain $= 3/2$.

where R_{OUT} is the equivalent output impedance of the switch array. Each gain configuration has a unique R_{OUT}, which is a function of the switching frequency, capacitor size and the switch impedance. Selecting a gain larger than G_{MIN} increases I_{MAX}. By increasing the gain only when needed, power is delivered more efficiently. The *gain-hopping* loop (Fig. 1) controls the gain based on a measure of the load current, and sets the

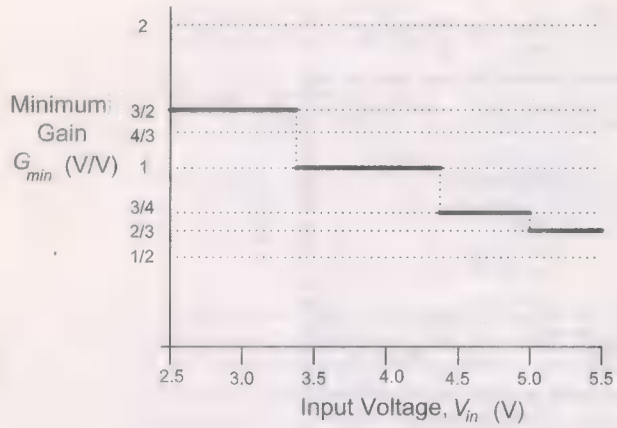


Fig. 4. G_{MIN} versus V_{IN} (for $V_{\mathrm{REG}} = 3.3$ V).

value of G_{MIN} as a function of V_{IN}. Fig. 4 illustrates the minimum gain versus V_{IN} for $V_{\mathrm{REG}} = 3.3$ V. The *gain-hopping* loop consists of an up-down counter, gain-set block, and a comparator. The up-down counter integrates the pulse sequence at the comparator output and directs the gain-set block to increase or decrease the gain.

The PFM loop in Fig. 1 contains a voltage reference V_{desired}, an analog comparator, and an oscillator. When V_{REG} is below the voltage reference, the switch array delivers current to the load. Alternately, when V_{REG} is above the reference, the switch array rests. By controlling the switching, the output impedance is modulated to provide the regulation. Also, for a given gain configuration, the pulse density of the comparator is proportional to I_{LOAD}. If I_{LOAD} is constant, the duty cycle of the output is fixed, resulting in a highly tonal frequency spectrum.

III. Modelling the Switched-Capacitor Regulator

In order to simulate the regulator at the system level, closed-loop expressions must be found for each of the gain configurations. That helps to predict the time-domain behavior of the regulator to a first-order approximation without simulating any real circuit components. The expressions that follow are all based on the assumption that the switches have zero on-resistance R_{on}. The output impedance of the regulator is a function of $R_{\mathrm{on}}, C_{\mathrm{ext}}$, and f (switching frequency). The assumption of R_{on} to be zero in the closed form expression predicts lower output impedance for the pump. This is similar to using a larger value of C_{ext} on the actual regulator.

A typical time-domain output of a given gain configuration ($G = 1/2$) is shown in Fig. 5. The two phases are $\Phi 1$ (gain phase) and $\Phi 2$ (common phase). The four voltages V_h, V_m, V_{m1}, and V_l at the boundaries of the two phases are of importance. Since a constant load I_{load} was assumed, the values of V_h, V_m, V_{m1}, and V_l repeat after every cycle in the steady state. By applying conservation of charge, one can compute the value of the output voltage V_m sampled at the end of phase $\Phi 2$ [3]:

$$V_m = \frac{V_{\mathrm{in}}}{2} - \frac{I_{\mathrm{load}}(C_{\mathrm{hold}} + C)}{8fC(2C + C_{\mathrm{hold}})}\left[1 + \frac{C_{\mathrm{hold}} + C}{3C + C_{\mathrm{hold}}}\right] - \frac{I_{\mathrm{load}}}{2f(3C + C_{\mathrm{hold}})} \quad (2)$$

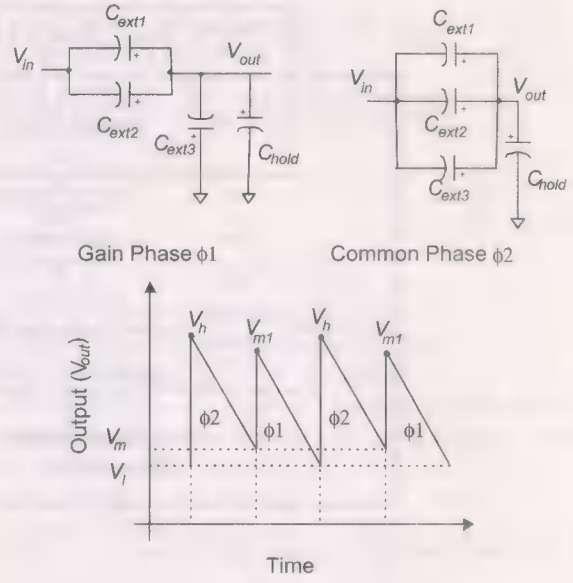


Fig. 5. Ideal time domain response for $G = 1/2$.

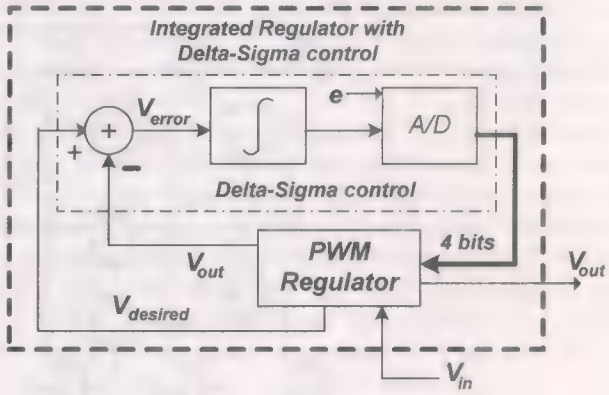


Fig. 6. Block diagram of the first-order $\Delta\Sigma$ control loop.

where f = switching frequency and C = $C_{\mathrm{ext1,2,3}}$, as all three capacitors are nominally of equal size. Clearly, if I_{load} is zero, the output voltage is $V_{\mathrm{in}}/2$, as expected. The above expression was simulated in MATLAB and compared with SPICE simulations. They were found to be in agreement.

One can also compute $V_m(n)$, the output voltage at the nth sample [3] for a time-varying input voltage $V_{\mathrm{in}}(n)$:

$$V_m(n) = aV_m(n-1) + bV_{\mathrm{in}}(n) \quad (3)$$

where

$$a = \frac{(C + C_{\mathrm{hold}})^2}{(3C + C_{\mathrm{hold}})^2}$$

and

$$b = \frac{C}{(3C + C_{\mathrm{hold}})}\left[1 + \frac{C_{\mathrm{hold}} + C}{3C + C_{\mathrm{hold}}}\right]. \quad (4)$$

This suggests that the charge pump can be modeled as a lossy integrator with a pole at $a < 1$ and constant gain b. It should be mentioned that this model represents the charge pump in a single gain setting and does not model the dynamic variations between the different gain settings. The key idea is to be able to simulate the regulator to a first-order approximation, and to

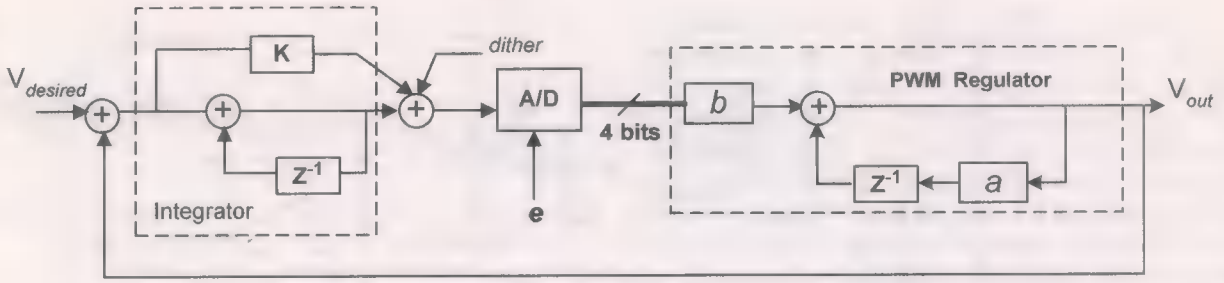


Fig. 7. Discrete time model of the regulator with the $\Delta\Sigma$ control loop.

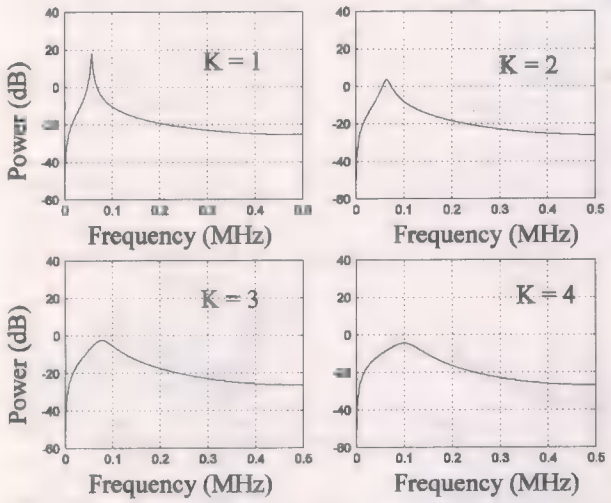


Fig. 8. Variation of the NTF with feedforward factor K.

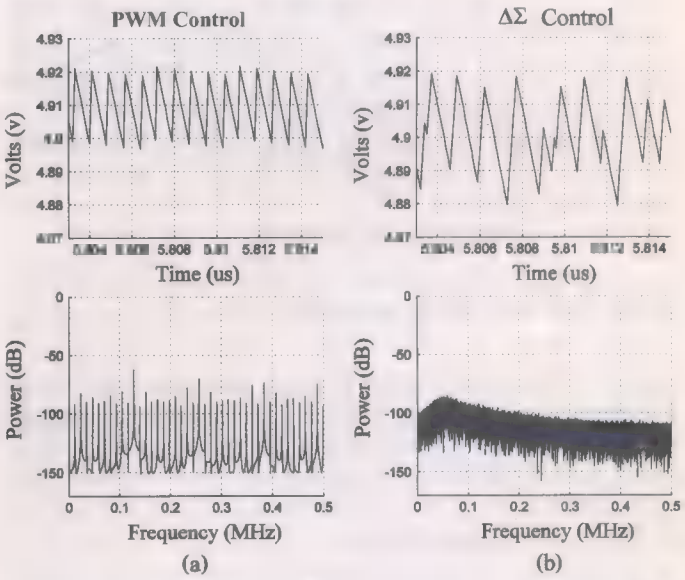


Fig. 9. Time and frequency-domain output plots for the regulator with and without the $\Delta\Sigma$ control loop.

predict the time- and frequency-domain responses without circuit-level simulation.

The efficiency of the charge pump can also be computed. The power dissipated at the output, $\mathrm{P_{out}}$, can be found, as we know V_{out} and I_{load}. To compute the power $\mathrm{P_{in}}$ supplied by the battery, we need to find the average current delivered by the input in each of the gain configurations. Then, the efficiency can be obtained from

$$\eta = \frac{P_{\mathrm{out}}}{P_{\mathrm{in}}} = \frac{V_{\mathrm{out}} I_{\mathrm{out}}}{V_{\mathrm{in}} I_{\mathrm{in}}}. \tag{5}$$

To calculate the average current I_{in} supplied by the input, we must find the charge supplied by V_{in} in every cycle. Since we know the value of V_{out} at the beginning and end of each clock phase [3], we can compute the amount of charge transferred and calculate the current supplied by V_{in} in every cycle. These computations do not take into account the nonzero switch resistance and the power dissipation in the other regulator circuits. The predicted efficiency given by the closed form expression will be close to the actual measured results. However, the closed form expression does not include the losses due to switching of parasitic capacitors associated with the big switches, nor the switching losses and I_q of the regulator. It is also inaccurate in the prediction of the efficiency when the regulator is hopping from one gain to another.

IV. Delta-Sigma Control Loop

As mentioned earlier, the burst-mode (PFM) control mechanism leads to a tonal spectrum for the output ripple, which may introduce excessive noise into the signal band of the device powered by the regulator. The tones may be converted into filtered pseudo-random noise by incorporating the complete regulator as the feedback DAC into a delta-sigma loop, as shown in Fig. 6. We assume that the quantization error $e[n]$ can be modeled as an additive white noise which is independent of the input, is uniformly distributed in $[-\Delta/2, \Delta/2]$ where Δ is the step size of the quantizer, and has a white power spectral density [4]. Then $e[n]$ can be represented as an additional input to the linearized system. The output of the modulator $\mathrm{Y}(z)$ can be expressed as

$$Y(z) = \mathrm{STF}(z)U(z) + \mathrm{NTF}(z)E(z) \tag{6}$$

where $\mathrm{STF}(z)$ is the signal transfer function, and $\mathrm{NTF}(z)$ is the noise transfer function. For the first-order $\Delta\Sigma$ modulator

$$\mathrm{STF}(z) = \left.\frac{Y(z)}{U(z)}\right|_{E(z)\equiv 0} = \frac{H(z)}{1+H(z)} \tag{7}$$

$$\mathrm{NTF}(z) = \left.\frac{Y(z)}{E(z)}\right|_{E(z)\equiv 0} = \frac{1}{1+H(z)}. \tag{8}$$

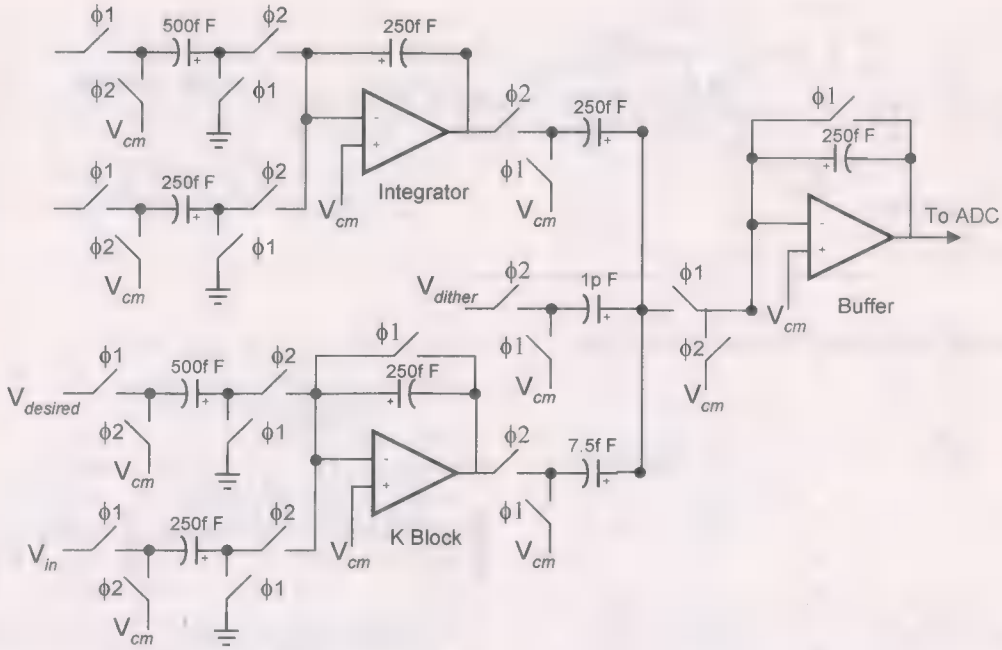


Fig. 10. Delta-sigma control implementation.

Equation (8) illustrates that if H(z) is a low-pass function with a high low-frequency gain, the quantization noise is high-pass filtered.

A. Delta-Sigma Control Loop

The simplified model of the modified regulator with a delta-sigma control loop is shown in Fig. 6. The $\Delta\Sigma$ loop provides a 3-bit word necessary for gain selection, plus the 1-bit *skip* signal for the PFM operation. The $\Delta\Sigma$ loop contains an integrator and a 4-bit analog-to-digital converter (ADC). The charge pump acts as the digital-to-analog converter (DAC) in the loop. The output of the DAC is the regulated voltage.

The error between the desired voltage and the output voltage is integrated and fed to the 4-bit ADC. As the output voltage approaches the desired voltage, the error signal decreases, reducing the input to the ADC. This causes a smaller gain to be chosen, until the minimum gain is reached. Since the $\Delta\Sigma$ control is a first-order loop, dither must be injected to avoid tone generation [5], [6].

The 3 MSBs from the A/D select one of the seven gain levels, and the LSB controls the PFM operation. Since there are seven possible gain settings, the 3 bits are sufficient to control all possible gains.

B. Discrete-Time Model of the Delta-Sigma Control Loop

Fig. 7 illustrates the discrete-time model of the $\Delta\Sigma$ control loop with the regulator. The delta-sigma loop is a first-order loop and by itself it is unconditionally stable. As mentioned earlier, the charge pump can be modeled as a lossy integrator which creates an additional pole and may make the loop unstable. In order to stabilize the loop, a feedforward path was added around

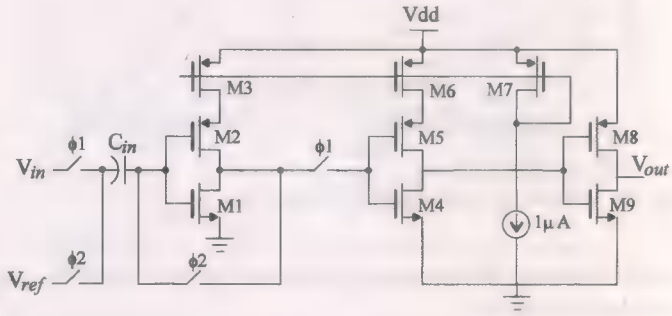


Fig. 11. Clocked CMOS comparator.

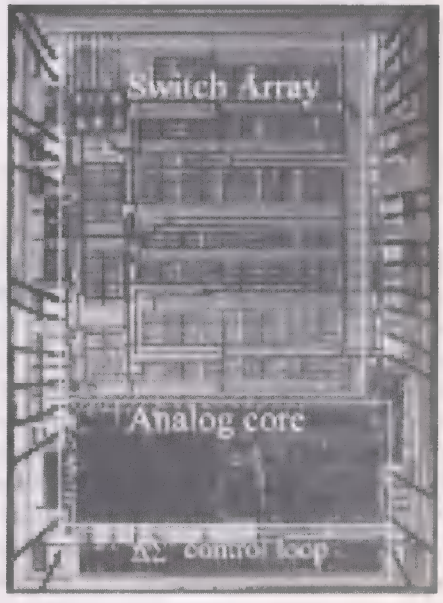


Fig. 12. Die photograph of regulator with $\Delta\Sigma$ control loop.

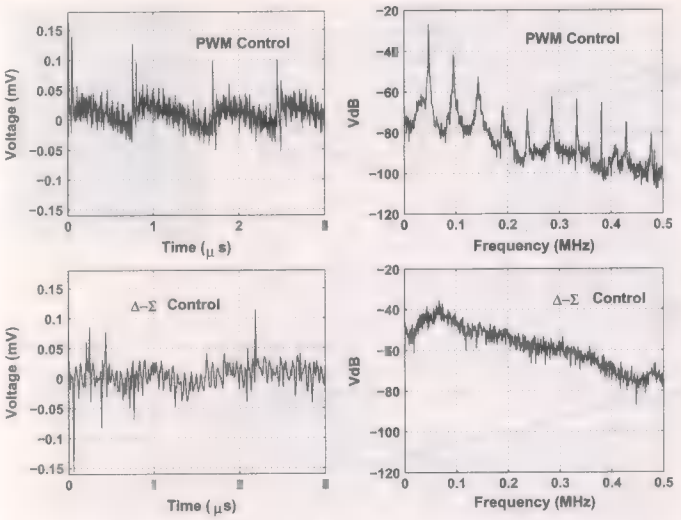


Fig. 13. Measured output ripple and output spectrum with PWM control and $\Delta\Sigma$ control for $I_{\text{load}} = 50$ mA, $V_{\text{out}} = 3.2$ V and $V_{\text{in}} = 3.7$ V.

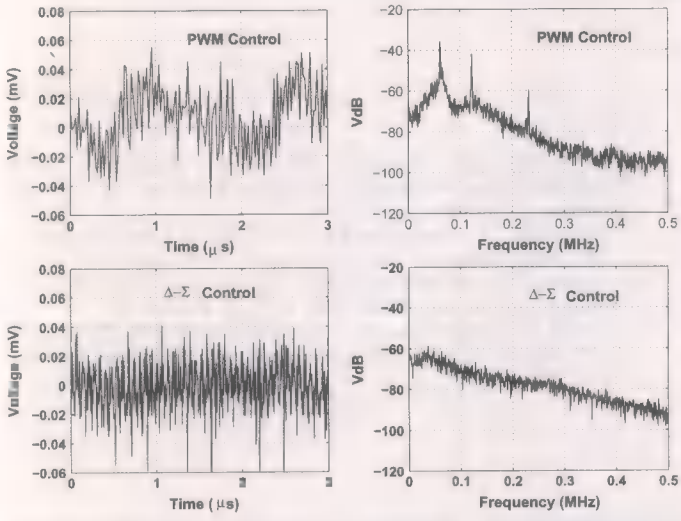


Fig. 14. Measured output ripple and output spectrum for PWM control and $\Delta\Sigma$ control loop for $I_{\text{load}} = 150$ mA, $V_{\text{out}} = 3.2$ V and $V_{\text{in}} = 3.7$ V.

the integrator with a gain K. The NTF for the system shown in Fig. 8 is given below:

$$\text{NTF}(z) = \frac{V_{\text{out}}}{E} = \frac{b(1-z^{-1})}{1 - z^{-1}[1 + a - (K+1)b] + z^{-2}(a - Kb)} \quad (9)$$

where E is the quantization error of the ADC. This is valid for a specific value of the input and output voltages and load current, and assumes that the system is settled. It does not represent the dynamic behavior of the system, but gives a good estimate of the stability of the system. We see peaking in the NTF which indicates some instability in the loop when the delta-sigma control is wrapped around the regulator.

The NTF is shown in Fig. 8 for different feedforward gains. As K increases, the pole-Q reduces, making the system more stable. This can be intuitively explained as the feedforward path reduces the effect of the delay through the integrator. We have not been able to come up with a closed form expression for stability for the entire system, but MATLAB simulations indicated that adding a feedforward reduces the peaking in the NTF, and a feedforward factor (K) greater than 4 does not benefit stability. The time-domain output and the output spectrum of the regulator with and without the $\Delta\Sigma$ loop are compared in Fig. 9. Both architectures were simulated using the closed-form equations [3] (corresponding to the time-domain response of Fig. 5). For the simulation C_{hold} was 30 μF, while $C_{\text{ext1,2,3}}$ was 0.33 μF and V_{in} was 5.2 V. The simulated curve matches closely the calculated NTF.

As Fig. 9 shows, $\Delta\Sigma$ control causes a slightly higher ripple. This can be attributed to the increased delay in the loop. However, the spectral properties are very much improved: instead of high-level tones, the output spectrum contains lower-level slightly colored noise, which is much less harmful in most applications.

V. Circuit Implementation

Since the $\Delta\Sigma$ loop (Fig. 6) controls only the gain selection, and is not a part of the signal path, it was kept very simple. The loop control circuitry is shown in Fig. 10. All the circuitry was single-ended since the LSB was large (150 mV). The integrator and the gain block were standard switched-capacitor stages. The unit capacitance used was 250 fF. A simple two-stage Miller-compensated operational amplifier, with an open-loop gain of 65 dB, a unity-gain frequency of 17 MHz and a phase margin of 55 degrees was used. The ADC/quantizer in the delta-sigma control loop was implemented as a conventional 4-bit flash structure [7].

A clocked CMOS comparator was used, as shown in Fig. 11. The LSB of the ADC is large, so an inverter based comparator could be used. The inverters contain current sources to limit the current flow and hence the power dissipation. A resistor ladder sets the reference voltage levels. The total resistance of the ladder is 220 kΩ. The dither circuit is a pseudo-random number generator using flip-flops and XOR gates. The voltage reference block consists of a bandgap reference, a D/A converter and an E^{2}PROM block. This generates the V_{desired} values ranging from 3 to 5 V. The E^{2}PROM allows post-package trimming of the bandgap voltage and V_{REG} adjustments through the DAC.

VI. Experimental Results

A prototype regulator incorporating the delta-sigma control loop was implemented in a 0.72-μm CMOS technology. The die photo is shown in Fig. 12. The active die area is 2.45×3.1 mm^2. The area of the control loop is 2.45 mm $\times$ 0.4 mm. The fabricated chip was tested through the input range of 3–5 V for several loads and output voltages. Typical measured output ripple and spectrum curves for load currents of 150 and 50 mA, an output voltage 4.7 V, and input voltage 3.4 V are shown in Figs. 13 and 14. The measurement bandwidth was 500 kHz. We can see that the PFM control has larger noise spikes at lighter loads and lesser spikes at heavier loads. This can be attributed to the fact that the PFM control "skips" less at higher loads. For

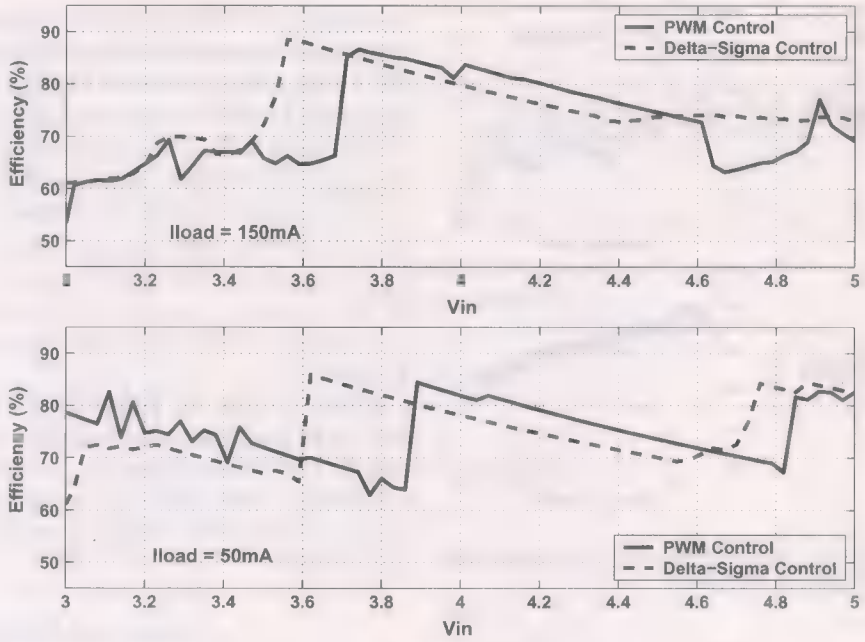


Fig. 15. Measured efficiencies for PWM control and Delta-Sigma control for $V_{\text{out}} = 3.2$ V and $V_{\text{in}} = 3.7$ V.

this reason the noise floor of the regulator with delta-sigma control is higher in the light loads then at heavier loads (as the total noise is not removed).

The efficiencies of the PFM and $\Delta\Sigma$ architectures are plotted in Fig. 15. With the delta-sigma control loop the efficiency curves are smoother than with the PFM control loop. The $\Delta\Sigma$ control loop selects a lower gain faster than a traditional PFM control loop. However, once the minimum gain has been chosen, the efficiencies are comparable for the two architectures.

VII. Conclusion

A pulse-frequency-modulation voltage regulator with a $\Delta\Sigma$ control loop was designed and fabricated. The test results indicate that the suppression of noise tones is possible using this technique. The additional delay through the loop increased the ripple and caused slightly poorer regulation, but gave much better spectral behavior.

Acknowledgment

The authors would like to thank the following people for their technical help and support: M. Fraley, R. Batten, B. Chatterjee, M. Keskin, J. Silva, J. Parry, T. Glad, S. Close, P. Wong, and R. Perigny.

References

[1] A. Rao, W. McIntyre, U. Moon, and G. Temes, "A noise-shaped switched-capacitor DC-DC voltage regulator," in *PROC. IEEE Eur. Solid-State Circuits Conf.*, Sep. 2002, pp. 375–378.

[2] J. Kotowski, W. J. McIntyre, and J. P. Parry, " Capacitor DC-DC converter with PFM and gain hopping," U.S. Patent 6,055,168, Apr. 25, 2000.

[3] A. Rao, "An efficient switched-capacitor buck-boost voltage regulator using delta-sigma control loop," M.S. thesis, Oregon State University, Corvallis, OR, May 2002.

[4] J. C. Candy and O. J. Benjamin, "The structure of quantization noise from delta-sigma modulation," *IEEE Trans. Commun.*, vol. COM-29, no. 9, pp. 1316–1323, Sep. 1981.

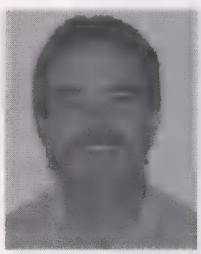
[5] V. Friedman, "Structure of the limit cycles in delta-sigma modulation," *IEEE Trans. Commun.*, vol. 36, no. 8, pp. 972–979, Aug. 1988.

[6] I. Galton, "One-bit dithering in delta-sigma modulator-based D/A conversion," in *Proc. IEEE ISCAS*, May 1993, pp. 1310–1313.

[7] A. G. F. Dingwall, "Monolithic expandable 6-bit 20-MHz CMOS/SOS A/D converter," *IEEE J. Solid-State Circuits*, vol. 14, no. 6, pp. 926–931, Dec. 1979.

Arun Rao received the B.S. degree from Bangalore University, India, and the M.S. degree from Oregon State University, Corvallis, in 1998 and 2002, respectively.

He worked as a Design Engineer in the Data Communication Group at Cypress Semiconductor Corporation from 1998 to 2000. Currently, he is a Design Engineer with National Semiconductor at the Grass Valley, CA, Design Center.



William McIntyre (M'90) received the B.S. and M.S. degrees in electrical engineering from the University of California at Davis in 1988 and 1990, respectively.

He was a Design Engineer in the Communications division for Intel Corporation from 1990 to 1993. From 1993 to 1995, he worked for Silicon Systems Inc. in the Communications and Industrial Products Division. Since then, he has been with National Semiconductor at the Grass Valley, CA, Design Center, where he is a Member of the Technical Staff, working on ICs for portable power applications. His research interests are in the area of switched capacitor dc–dc converters including regulation methods and switch arrays. He holds six patents.

Un-Ku Moon (S'92–M'94–SM'99) received the B.S. degree from the University of Washington, Seattle, the M.Eng. degree from Cornell University, Ithaca, NY, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, all in electrical engineering, in 1987, 1989, and 1994, respectively.

From February 1988 to August 1989, he was a Member of Technical Staff at AT&T Bell Laboratories, Reading, PA, and during his stay at the University of Illinois at Urbana-Champaign, he taught a microelectronics course from August 1992 to December 1993. From February 1994 to January 1998, he was a Member of Technical Staff at Lucent Technologies Bell Laboratories, Allentown, PA. Since January 1998, he has been with Oregon State University, Corvallis. His interests have been in the area of analog and mixed analog–digital integrated circuits. His past work includes highly linear and tunable continuous-time filters, telecommunication circuits including timing recovery and analog-to-digital converters, and switched-capacitor circuits.

Prof. Moon is a recipient of the National Science Foundation CAREER Award in 2002, and the Engelbrecht Young Faculty Award from Oregon State University College of Engineering in 2002. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART II: ANALOG AND DIGITAL SIGNAL PROCESSING. He also serves as a member of the IEEE Custom Integrated Circuits Conference Technical Program Committee and the Analog Signal Processing Program Committee of the IEEE International Symposium on Circuits and Systems.

Gabor C. Temes (LF'98) received his undergraduate education at the Technical University and Eotvos University, Budapest, Hungary, and the Ph.D. degree in electrical engineering from the University of Ottawa, Canada, in 1961. He received an honorary doctorate from the Technical University of Budapest in 1991.

He was on the faculty of the Technical University of Budapest from 1952 to 1956. He worked as a Project Engineer at Measurement Engineering Ltd., Arnprior, Canada, from 1957 to 1959. From 1959 to 1964, he was a Laboratory Supervisor at Northern Electric R&D Laboratories (now Bell-Northern Research), Ottawa, Canada. From 1964 to 1966, he was a Research Group Leader at Stanford University, Stanford, CA, and from 1966 to 1969, he was a Corporate Consultant at Ampex Corporation, Redwood City, CA. Between 1969 and 1991, he was on the faculty of the University of California of Los Angeles (UCLA). He is now an Emeritus Professor at UCLA and a Professor in the Department of Electrical and Computer Engineering at Oregon State University (OSU), Corvallis. He has served as Department Head at both UCLA and OSU. He is co-editor and co-author of *Modern Filter Theory and Design* (Wiley, 1973), co-author of *Introduction to Circuit Synthesis and Design* (McGraw-Hill, 1977), co-author of *Analog MOS Integrated Circuits for Signal Processing* (Wiley, 1986), and co-editor and co-author of *Oversampling Delta-Sigma Data Converters* (IEEE Press, 1992) and *Delta-Sigma Data Converters* (IEEE Press, 1997), as well as a contributor to several other edited volumes. He has published approximately 300 papers in engineering journals and conference proceedings. His recent research has dealt with CMOS analog integrated circuits, as well as data converters and integrated interfaces for sensors.

Dr. Temes was an Associate Editor of the *Journal of the Franklin Institute*, Editor of the IEEE TRANSACTIONS ON CIRCUIT THEORY, and Vice President of the IEEE Circuits and Systems Society. In 1968 and in 1981, he was co-winner of the Darlington Award of the IEEE Circuits and Systems Society. In 1981, he received the Outstanding Engineer Merit Award of the Institute for the Advancement of Engineering. In 1982, he won the Western Electric Fund Award of the American Society for Engineering Education, and in 1984 received the Centennial Medal of the IEEE. He received the Andrew Chi Prize Award of the IEEE Instrumentation and Measurement Society in 1985, the Education Award of the IEEE Circuits and Systems Society in 1987, and the Technical Achievement Award of the same Society in 1989. He is the recipient of the 1998 IEEE Graduate Teaching Award and the IEEE Millennium Medal, as well as the IEEE CAS Golden Jubilee Medal in 2000.

A 126-μW Cochlear Chip for a Totally Implantable System

Julius Georgiou, *Member, IEEE,* and Christopher Toumazou, *Fellow, IEEE*

***Abstract*—In this paper, a single-chip speech processor/stimulator is presented for use in a totally implanted cochlear prosthesis system. It implements a continuous interleaved sampling (CIS) strategy. By combining the speech processor and the stimulator into one mixed-signal chip, both size and power are reduced sufficiently, so as to make a totally implanted system feasible. First silicon has been validated and typically operates at 126 μW (excluding cochlear stimulation currents).**

***Index Terms*—Analog signal processing, cochlear implant, micropower, subthreshold.**

I. Introduction

THE worldwide deaf population exceeds 70 million, of which approximately 600 000 profoundly deaf individuals are found in the US and 420 000 in the UK. Although conventional hearing aids provide considerable help for the majority of individuals with mild, moderate, or severe hearing loss, these aids are of little help where the deafness is *profound* (average loss is greater than about 90 dB SPL in both ears). In such cases, an invasive electronic device, i.e., a cochlear implant, has the capability to restore hearing to some degree. A cochlear implant is used to replace the damaged natural hearing components from the eardrum up to the inner hair cells, which transduce fluid motion into electrical signals in the nerves. In general, a cochlear implant consists of an external speech processor and an implanted receiver stimulator; the speech processor picks up audio signals and processes these in a suitable manner, so as to maximize the benefit for each particular patient. The processed signal is then transmitted to the implanted receiver, which produces charge-balanced electrical signals to stimulate the auditory nerve. This gives a degree of hearing sensation and prevents further nerve degeneration [1].

Current cochlear speech processors, regardless of manufacturer, are heavily based upon digital technology running DSP algorithms on ASIC processors. Although digital technology has the advantage of being more flexible to modifications through software, there is a high power penalty to be paid when the required precision is below 8 bits [2]. As the electrical dynamic range of patients' remaining neurons range between 3 and 20 dB, using more than 8-bits precision for the signal processing is a massive overkill. With the best state-of-the-art digital speech processors, batteries need changing every day or two, and most patients, given the choice, would prefer not to wear an externally visible processor, although "behind-the-ear" (BTE) systems have recently reached the market (Fig. 1.). Prior work, to move away from the digital trend and return to low-power analog subthreshold systems, has either solved only a small part of the problem [3] or not aimed at the application of cochlear implants but at modeling the basilar membrane [4]–[6].

Fig. 1. Illustration of a digital-processor-based state-of-the-art cochlear implant system.

By adopting the best of both the digital and analog worlds, a complete system, with sufficiently low power consumption to be totally implanted, is presented; digital circuitry is used for robust communication with the implant, primarily for control purposes, while low-power analog circuits are used for the signal processing.

A totally implantable system is desired by manufacturers and patients alike for the following reasons.

Improved Aesthetics: A totally concealed cochlear prosthesis can bring significant improvements in self-confidence and third-party attitudes, as has been witnessed with "in-the-canal" hearing aids. Blending in within mainstream educational institutions becomes significantly easier for children.

Reduction of Practical Limitations: A totally implanted system will allow the recipients to engage in activities they were otherwise unable to do while maintaining hearing, e.g., swimming, water-skiing, windsurfing, etc.

Improved Perception: By having the microphone implanted in the canal, the patient can make use of the directional amplification provided by the external pinna, while also reducing noise from wind, an effect observed from "in-the-canal" hearing aids. The removal of the data rate restriction between the implanted part and the external processor allows the use of a higher temporal resolution, without compromising the number of active channels. The positive effect on patient speech recognition of increased temporal and frequency resolution is well known [7].

Manuscript received March 2, 2004; revised August 13, 2004. This work was supported in part by EPSRC under Grant GR/R96583/01, Epic Biosonics Inc. (Canada), Toumaz Technology Ltd. (U.K.), and Geosilicon Ltd. (Cyprus).

J. Georgiou was with the Electrical Engineering Department, Imperial College, London, U.K. He is now with Geosilicon Ltd., Nicosia 2020, Cyprus (e-mail: julio@geosilicon.com).

C. Toumazou is with the Department of Electronic and Electrical Engineering, Imperial College, London WC1E 7JE, U.K. (e-mail: c.toumazou@imperial.ac.uk).

Digital Object Identifier 10.1109/JSSC.2004.840959

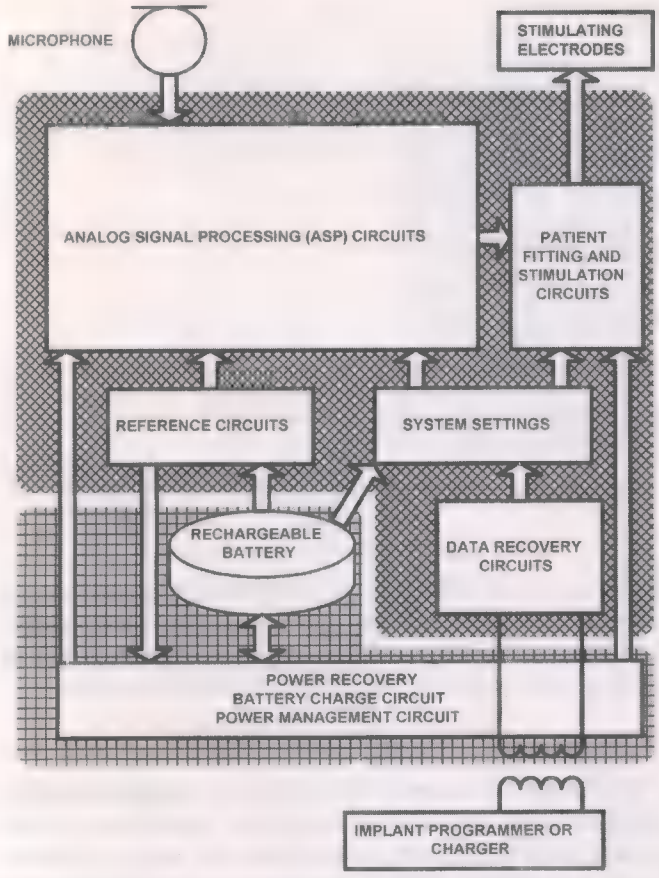


Fig. 2. Block diagram of described cochlear implant system.

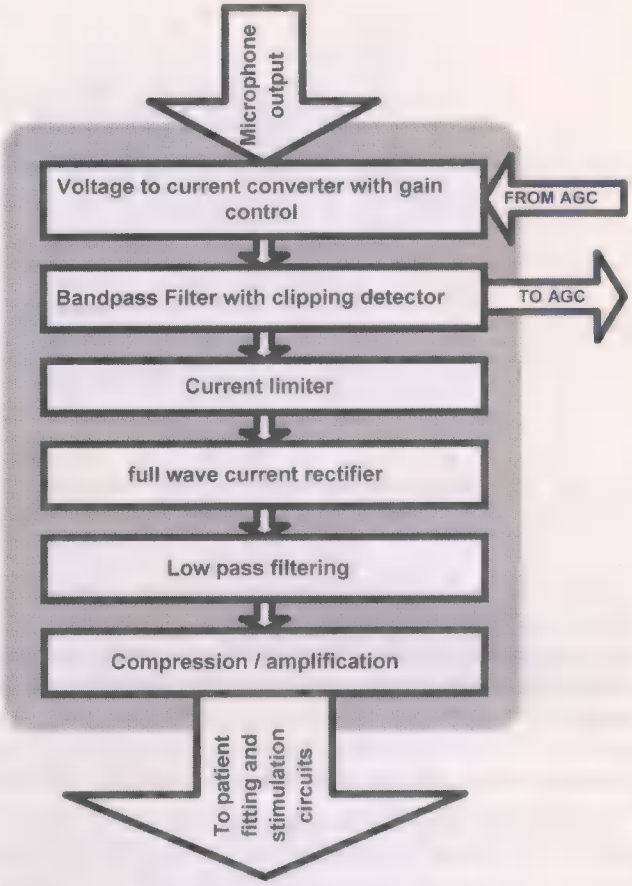


Fig. 3. Analog functions in a single channel.

II. System Overview

The system consists of a single chip that combines the audio processing/stimulation circuits, a rechargeable battery, and a second chip containing power management and charging circuits. All system components are encapsulated in a hermetically sealed platinum case for biocompatibility reasons. A block diagram of the complete system is shown in Fig. 2. Power and system settings from the outside world are transferred to the implant via an inductive link, using a PWM scheme by means of an implant programmer or charger.

The viability of such a prosthesis can be attributed to novel electrode designs that reach closer to the auditory nerve endings in the cochlea; thus, the overwhelmingly power-hungry stimuli of the past have been reduced to consume power comparable to, or less than, that used by the speech processor. In addition, the sufficient maturing of the cochlear implant speech processing algorithms has made the complete reprogrammability of DSPs unnecessary.

This paper will only detail the components of the audio processing/stimulation chip. This chip (diagonal cross-hatching) has been manufactured in a 0.8-μm (5 V) process with direct portability to a 0.8-μm, high-voltage (5 and 20 V) process. This option is necessary because the upper voltage needed for stimulation is reviewed as electrode technology develops; the upper voltage is simply a function of the maximum comfortable stimulation current and the maximum cochlear-electrode impedance at this current. The impedance can be influenced by how close the electrodes get to the neurons, by the materials used, and by the surface area. These factors determine the upper voltage, and can only finally be determined *after* clinical trials of novel electrodes.

III. Analog Signal Processing

A. Underlying Technology

Given the constraint that the voltage of the system is to be kept at no less than 4 V for stimulation purposes, reducing power implies reducing current levels. Hence, the system was designed to operate predominantly in the subthreshold region (FET technology was necessary for high integration density of the digital control and trimming circuits). In the past, the subthreshold region has been avoided, as device models were poor, and device matching even poorer [8]; the EKV and the BSIM (v3.3 onwards) models can currently cope quite well with the continuous modeling of the all the FET operating regions. Similarly, as the feature sizes have been reduced, the quality of the gate oxide has improved such that, per unit square gate area, matching has also improved [9]. In terms of dynamic range, the subthreshold region usually can provide around 60 dBs if carefully designed.

B. Stimulation Strategy

As analog systems are not as easily reconfigurable as digital systems, the choice of processor stimulation strategy is critical in making a successful implant system. Various studies have

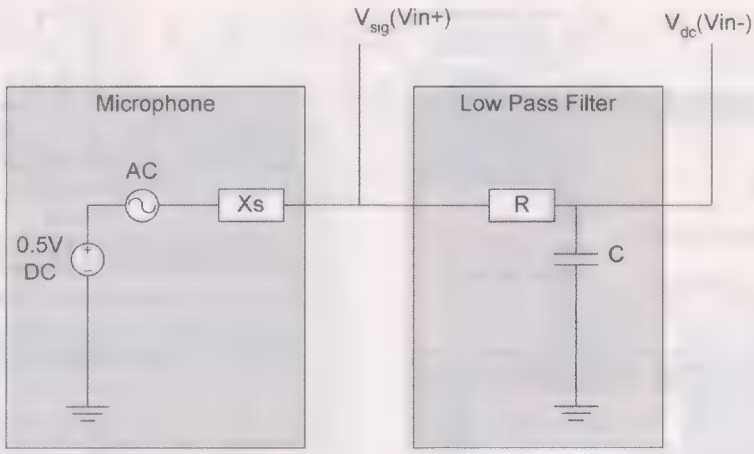


Fig. 4. An offchip *RC* low-pass filter creates a high-pass function when driving a differential input and also produces good single-to-double-ended conversion. Xs is the output impedance of the electret microphone that is approximately 4 kΩ.

shown that the performance of fast continuous interleaved sampling (CIS) strategies provide better results in comparison to other strategies [7], [10], [11], especially those that attempted to preprocess speech and extract particular characteristics to present to the brain.

C. Analog Signal Processor Overview

The analog signal processor consists of two sets of eight parallel channels, whose center frequencies are logarithmically distributed in a fashion similar to that employed in the natural cochlea. Fig. 3 shows the constituent functions through a single channel. The microphone's output is fed into a voltage-to-current converter (shared by a set of eight channels), which feeds the current-mode bandpass filter. The bias current of voltage-to-current converter is also used to adjust the input sensitivity of the system. An automatic gain control (AGC) circuit regulates a current so as to fit the largest audio signals into the 50-dB worst-case dynamic range of the filters. Each filter has an ultra-low-power clipping detector that consumes a maximum of 80 nW, given a 4-V supply. Each clipping detector output is fed to a common AGC circuit that reduces the voltage-to-current gain if clipping occurs in any of the channels. The attack and release times of the AGC are programmable and, if required, the AGC can be turned off when manual settings are preferred. A combined current-limiter/full-wave rectifier function block is placed in each channel after the filter. The current limiter is necessary in order to cut off large transients that may grow faster than the AGC's response time, hence, protecting the patient from uncomfortably large stimulation current pulses. The full wave rectifier is necessary for extracting the power in a particular audio band. Finally, a combined low-pass filter/compressor/current amplifier stage smoothes out the fully rectified signal and compresses it, such that uniform increments in sound levels are perceived accordingly by the patient, while also amplifying the signal from nanoampere current levels to the microampere levels needed for electrical stimulation.

The signal of each channel is then passed on to the patient fitting and stimulation circuits, which maximize a particular patient's comfort and hearing ability, and ensure that only one channel's signal is stimulating neurons at any one time according to the CIS strategy. Considerable power savings have been achieved by merging blocks and by using inherent functions provided by analog components. This will become more apparent when the individual circuits are presented.

D. Input Stage

1) Circuit Description: The electret microphone deemed suitable for this application can roughly be modeled as shown in the left half of Fig. 4, with Xs being the series output impedance of the microphone; the ac audio signal is superimposed on a 0.5-V dc signal. An off-chip *RC* low-pass filter is used to bias up the differential input to the system and also is used, in conjunction with the differential input, to create a high-pass filter that will attenuate 50/60 Hz mains pickup.

The voltage-to-current converter (Fig. 5) allows the transconductance to be tuned while still maintaining the same output dc current level maintained identically to the filter bias currents. Variations in the microphone's dc output level are easily tolerated with this circuit. Large device areas have been used in order to bring the flicker noise levels down sufficiently, and to provide reasonable matching; (1) and (2) model the drain current standard deviation and the flicker noise power, respectively.

$$\sigma\left(\frac{\Delta I_D}{I_D}\right) = \frac{A_{Idx}}{\sqrt{WL}} \tag{1}$$

where A_{Idx} is an empirical constant supplied for various values of overdrive voltage, i.e., $V_G - V_{T0}$.

$$\overline{I_{\text{flicker}}{}^2} \approx \frac{K I_{\text{sat}}^p \Delta f}{WL} \frac{1}{f}. \tag{2}$$

where K and p are process-dependant constants, WL the active area of the device, and Δf and f are the bandwidth and frequency, respectively.

The transconducting FETs' aspect ratios were kept such that they were well within the subthreshold region to maximize the efficiency, i.e., the gm/I ratio. The aspect ratio of current mirrors was lowered so as to improve current matching for a given current, i.e., by minimizing the coefficient A_{Idx} in (1). The device sizes are shown in Table I.

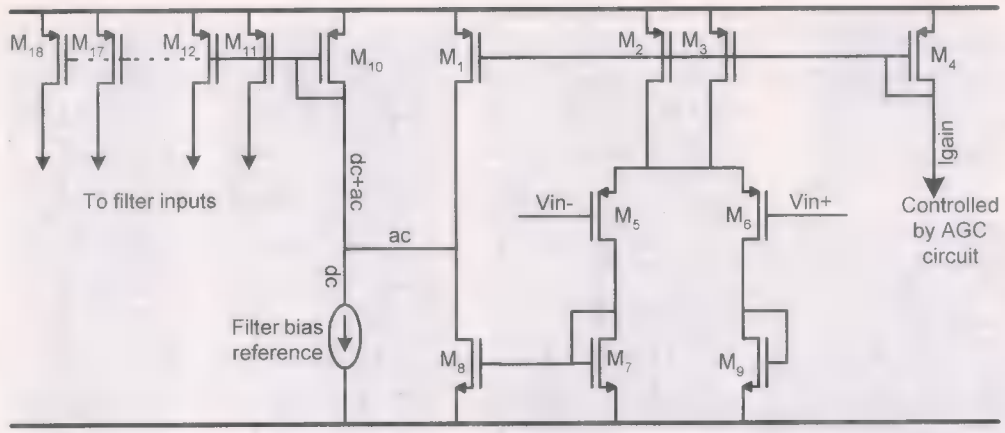


Fig. 5. Input transconductor.

TABLE I
DEVICE SIZES FOR INPUT STAGE

Device	W	L
M1-M4	80	60
M5-M6	240	5
M7-M9	50	120
M10-M22	10	10

The system has two independent front-end transconductors, each driving a bank of eight logarithmically spaced log domain filters. The system was split into the two different bias schemes as a method of pre-emphasis and dynamic range extension of the higher frequency bands, which have a relatively low energy content in speech. The frequency divide between vowels and consonants is generally found to be at about 1.2 kHz [7]. The higher bias current of the upper filter bank and its independent AGC allow for a better signal-to-noise ratio of the higher frequencies.

The input stage can be the most power-hungry part of the analog signal processing blocks, depending on the settings of the AGC. The current I_{gain} (see Fig. 5) that controls the transconductance varies from 10 to 200 nA.

2) Circuit Performance: At any instant, the dynamic range of the input stage, i.e., between the largest signal that will saturate the following filter and the input stage's noise floor, is on average 45 dB. This "capture window" is moved up or down with use of the AGC circuits, which can shift it over 30 dBs for the lower eight frequencies and 14 dBs for top eight frequencies. So the covered audio range for the upper eight frequencies is about 59 dB, while the covered audio range for the lower eight frequencies is about 75 dB. The worst-case total harmonic distortion (THD) figure was measured to be 3.8%, with the input stage at its minimum bias and an input signal corresponding to 91 dB SPL, which is very loud. More detailed results on the input stages THD can be found in reference [12]. Monte Carlo simulations predicted that over 99.7% yields should be expected, though actual circuit measurements showed that the Monte Carlo simulations to be more pessimistic than necessary.

E. Filters

1) Filter Design Description: A fully differential scheme for the filters was avoided, as subthreshold device matching in the 0.8-μm technology used was not sufficiently good to justify such a scheme. However, special care was taken to ensure substrate noise generated by digital circuits was sufficiently low and appropriately isolated; the ground guard separating digital and analog circuitry was 700 μm wide and had four bond wires attached to provide a low-impedance path for stray substrate noise. This solution was opted for, instead of a twin chip solution requiring chip-to-chip bonding, as the latter would be spatially wasteful within the package and was likely to reduce reliability. The extra silicon area used is not particularly important as production numbers are low and the chip costs are negligible in comparison to the complete product costs.

The filter used in this system is a derivative of one of the early log-domain filters [13]; log-domain filters [13]–[17] are linear when examined at a top level, however, no attempt is made to linearize the internal building blocks[1]. Benefits of such methodologies are that the circuits are not limited to small-signal operation; in addition, they generally have fewer constituent elements and can push a particular technology further in terms of frequency and lower voltage. When log-domain filters were originally conceived, they were designed with high-frequency/high-power operation in mind. However, with the exploitation of the subthreshold exponential characteristics [15], [19] for low power, these techniques were applied to audio frequency applications where device bandwidths in subthreshold can be an issue. Nevertheless, problems have been found in low-frequency weak-inversion implementations of log-domain filters [20], [21]; these problems are related to the presence of multiple dc operating points that are not present in the bipolar versions, ultimately because the bipolar devices have a smaller "triode" operating region. A method for eliminating the unwanted operating points with marginal additional power consumption was developed by the authors to overcome this problem [22] for the cochlea prosthesis. The circuit diagram of the single operating point implementation is shown in Fig. 6. A signal is input to the filter via device M_3. The current mode signal is transformed into the voltage log-domain via device M_{12} (All the remaining current sources are biased. with a constant current.) At nodes v_1 and v_2, the nonlinear positive output conductance of the E+ cells is can-

[1]The most basic building block element is the exponential, inherent from the voltage to current characteristics of bipolar transistors.

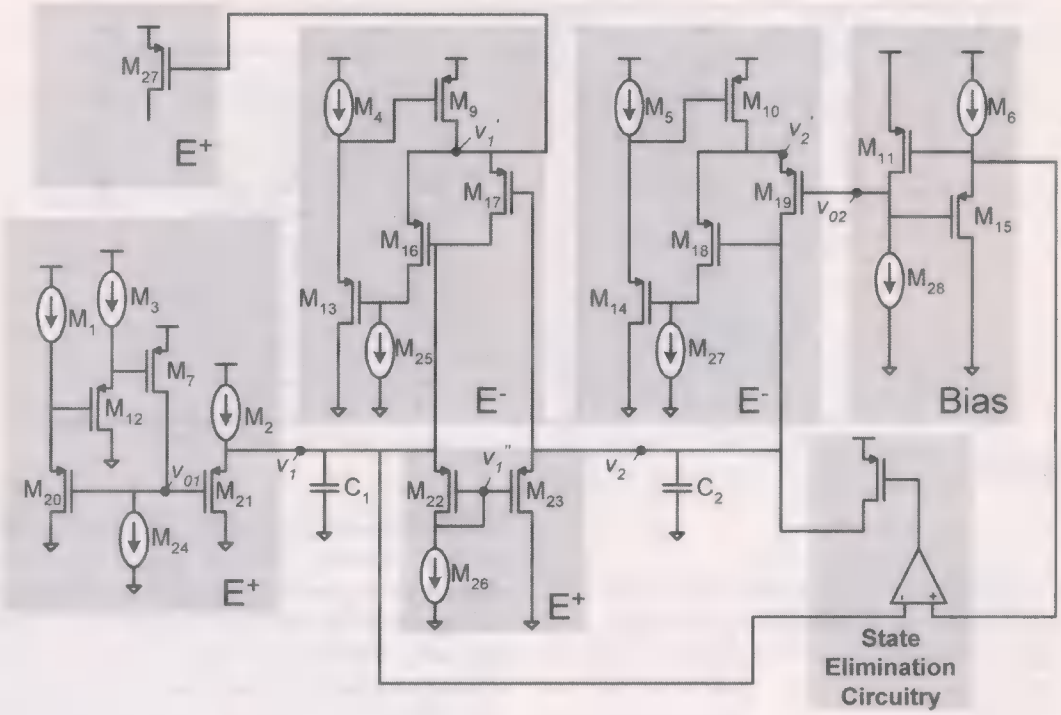


Fig. 6. Filter circuit schematic. All devices are sized 10 μm$\times$ 10 μm. All current values are set at I_0, except those of M_2 and M_1, which control the filter's Q and which are set at $I_0 + I_0/Q$. The input is put through M_3 and the bandpass output taken via M_{27}.

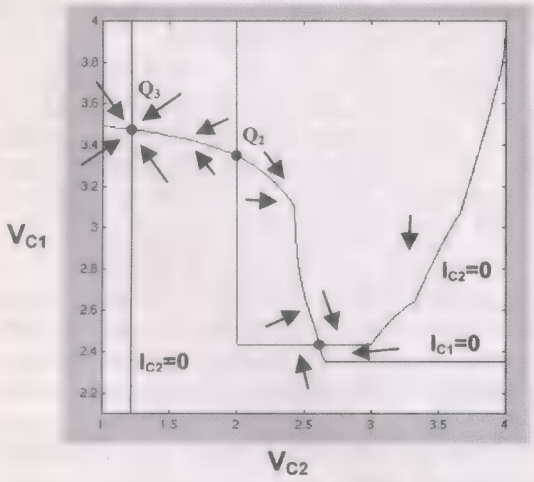


Fig. 7. Phase portrait of the original subthreshold filter's operating states [20], [21]. The desired dc operating point is Q_1. The undesired stable dc operating point is Q_3.

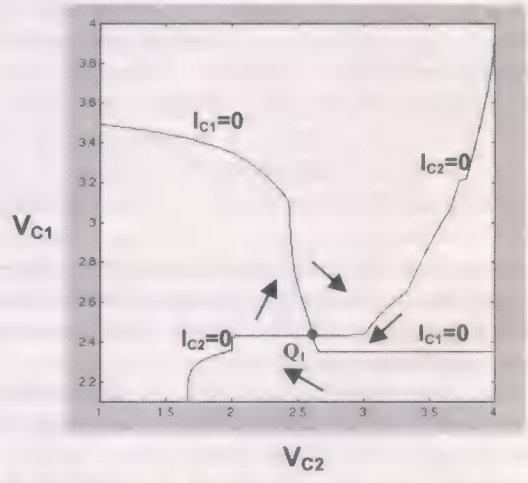


Fig. 8. Phase portrait of the corrected subthreshold filter's operating states. There is only one operating point, i.e., Q_1.

celled or reduced by the nonlinear negative output conductance of the E$-$ cells. The Q of the filter is therefore controlled by bias current M_2, which regulates the output conductance of the input E+ cell and provides for damping. The filtered output is provided by device M_{27} that expands the level shifted voltage v_1' back into the linear current domain. The state elimination circuitry keeps the associated pMOS device off during normal operation, during which there is one V_{gs} drop across the positive and negative input terminals of the comparator. The negative voltage terminal exceeds that of the positive when approaching the unwanted quiescent operating point, so current is sunk into node v_2 to keep operating in the desired region.

The circuit's operating points are found by replacing capacitors C_1 and C_2 with voltage sources V_1 and V_2, while sweeping every combination of voltages, during which currents I_1 and I_2 flowing through the voltage sources V_1 and V_2 are monitored. By using the *contour* function in MATLAB, one can plot the extrapolated zero current contour for $I_1 = 0$ and for $I_2 = 0$. Where the two contours meet is an indication of a dc operating point, since when there is no current driven into the capacitors, the voltages of each capacitor will remain at the same level. The direction of the state space trajectories can be obtained by using the *quiver* command whose components are determined by I_1 and I_2. Fig. 7 shows the filter's operating points without the state elimination circuitry. Fig. 8 shows the single operating

TABLE II
SIMULATED FILTERS THD AT 50% AND AT 95% MODULATION INDEX

I_{bias} /nA	Filter no	Frequency /Hz	Modulation Index	Simulated THD /%
10	1	300	50%	0.24
10	1	300	95%	3.7
10	8	1250	50%	0.9
10	8	1250	95%	3.9
50	9	1580	50%	0.67
50	9	1580	95%	4.4
50	16	6450	50%	1.3
50	16	6450	95%	4.9

TABLE III
MEASURED COMBINED INPUT-STAGE AND FILTERS THD AT 90% MODULATION INDEX

I_{bias}	I_{gain}	Filter	F_0/Hz	Measured THD
10	10	2	365	3.8%
10	10	4	565	3.9%
10	10	8	1260	3.8%
50	50	9	1580	4.2%
50	50	12	2810	4.5%
50	50	16	6300	6%

point achieved once the elimination circuitry is added. More details concerning the filter and its stability analysis can be found in [12].

The circuit was implemented using solely pMOS devices (apart from a few nMOS bias current sinks that do not require exponential operation) for a number of reasons. First, the fact that the pMOS devices have their own well means that the bulk-source voltage can be set to zero, thus simplifying the weak-inversion drain-source current expression. The well also provides some protection against substrate noise. Second, pMOS devices are less noisy and have better current matching than their nMOS counterparts in the technology used.

The filter's center frequency f_0 is given by

$$f_0 = \frac{I_0}{2\pi n \phi_t C} \tag{3}$$

where n is the subthreshold parameter and ϕ_t is the thermal voltage. This expression gives the designer the choice of determining the center frequencies using the bias current I_0 or the capacitance C. As the weak inversion region has a somewhat limited dynamic range, the logarithmic spacing of the eight filters in each of the two filterbanks was determined by adjusting the capacitor sizes and keeping the bias current consistent so as to keep the devices in the optimal subthreshold operating point, hence, keeping signals well above the noise/leakage levels and below the moderate inversion region. Post-implant fine-tuning, if required, can be achieved by adjusting I_0 so as to maximize the hearing benefit in cases where the electrodes are inadequately inserted due to ossification of the cochlea.

2) Filter Performance: The aim of the two front-end transconductors and the eight filters connected to each of them is to separate the audio signal into its logarithmically distributed frequency bands. Distortion introduced after this separation has been conducted is insignificant, so long as the power content in the band remains the same. Hence, it makes sense to provide measured performance characteristics of the input stage and filter working together. Table II and Table III provide simulated and measured composite THD figures at the worst-case input bias situation spanning across both filter banks. These figures are typical of what can be measured for an audio input at around 91 dB SPL. With smaller sound signals, hence higher I_{gain} bias current, the input stage's linearity improves. It should be noted that measuring such small signals in current mode is not straightforward, as commercially available instruments are voltage mode. Table IV provides simulated THD results for just the filters. In evaluating these figures for the application, it is important to have in mind the relative crudeness of the electrodes. A significant amount of current spread is inevitable since the electrodes are bathing in an ionic fluid. Effectively, this causes some electrodes to activate neurons that are meant to be stimulated by their neighboring one.

TABLE IV
SIMULATED FILTER INPUT-REFERRED NOISE AND DYMAMIC RANGE

I_{bias}	Frequency	Filter	Input referred inband rms noise	DR
10	300Hz	1	2.8pA	57dB
10	1.25kHz	8	5.6pA	51dB
50	1.58KHz	9	13.3pA	58dB
50	6.45kHz	16	28pA	51dB

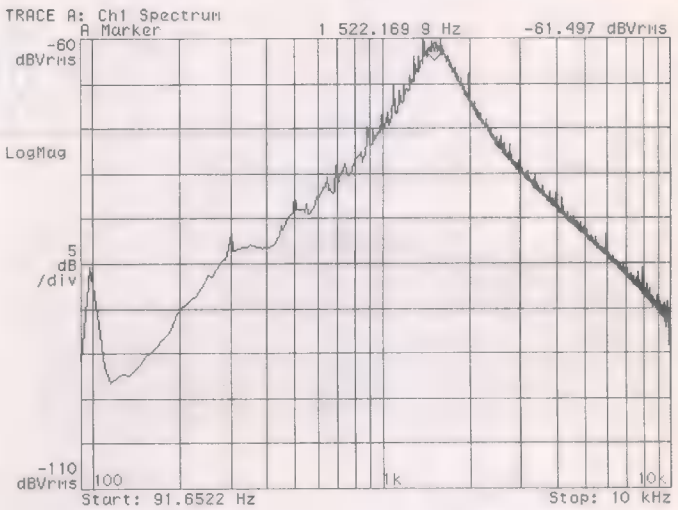


Fig. 9. Frequency response of the ninth filter. The measurement was made using custom-made V–I and I–V converters. Mains noise harmonics is an additional problem in making such measurements.

Depending on the particular filter, the simulated dynamic range varies between 51 and 58 dB. In practice (see Fig. 9) verifying these to be the filter's actual dynamic range was difficult as the custom-made I–V converter at the output was not suitable for measuring ac signals below 100 pA. Given that in Fig. 9 the maximum signal was around 10 nA, the minimum signal detected is around 100 pA. In addition, mains noise and harmonics were difficult to eliminate when measuring such small signal levels.

Clipping detection is achieved by making an extra copy of the output current and sinking it into a current source of twice the bias current. If the output exceeds this value, the current that

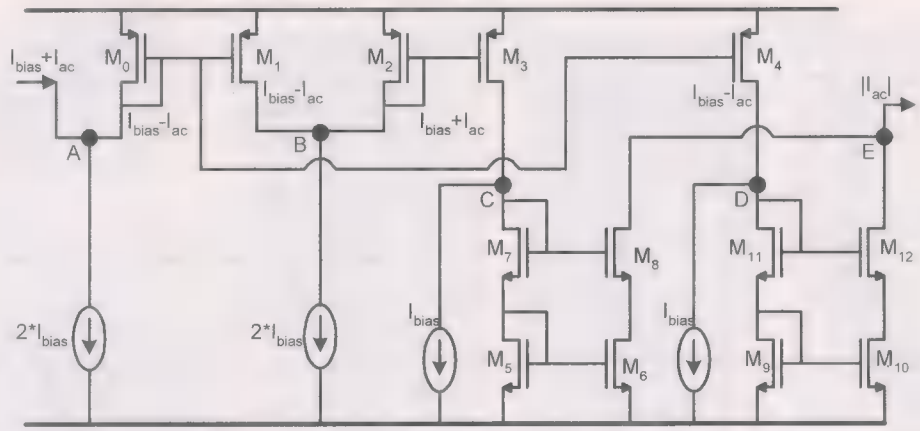


Fig. 10. Full-wave current rectifier and limitier circuit.

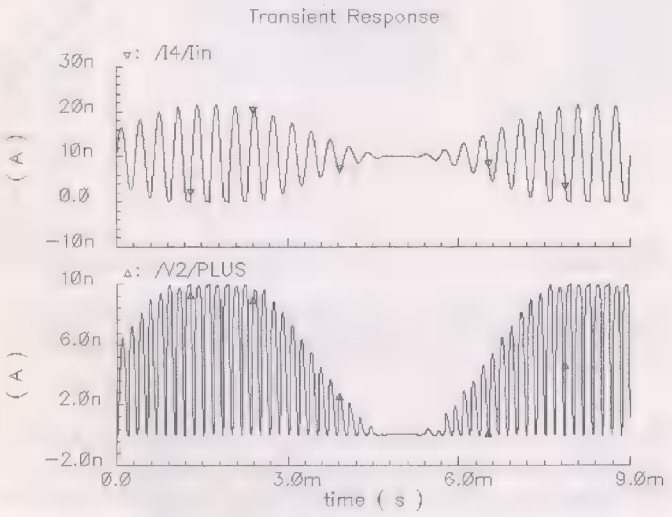


Fig. 11. A large input signal illustrates the current limiting feature of the current-mode ac signal full-wave rectifier.

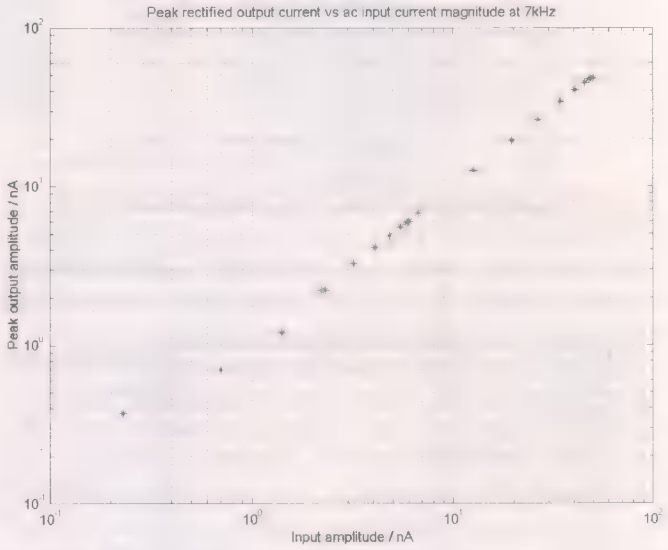


Fig. 12. Measured peak rectified output current versus peak ac input current at 7 kHz.

the source cannot sink will drive the node to the positive supply level, hence sending a signal to the AGC circuit to reduce gain if possible. The worst-case peak power consumption is of the order of 80 nW for a 4-V supply while it is nominally 40 nW with no input signal.

F. Current Limiter/Full-Wave Rectifier

The next two blocks in the signal processing chain shown in Fig. 3 will be dealt with together. The bandpass filter's output signal is Class A, i.e., an ac signal mounted on a dc bias. Since the filters are single-ended structures, in order to perform full-wave rectification it is necessary that both phases are recovered. Fig. 10 shows a schematic of the circuit which produces a full wave rectified copy of the filtered signal, and hard limits the output current to the twice the value of the bias current. The output current $I_{\rm bias} + I_{\rm ac}$ from the filter sourced to a node A, which is also connected to the input of current mirror ($\rm M_0$–$\rm M_1$) and to a current source drawing $2I_{\rm bias}$. From *KCL* it is apparent that the current drawn through the current mirror is $I_{\rm bias} - I_{\rm ac}$. The maximum current that can be drawn through this branch is limited to $2I_{\rm bias}$ given that the output of the filter can only provide unidirectional current. Similarly, the current $I_{\rm bias} - I_{\rm ac}$ is then sourced to node B onto which is connected to a current source drawing $2I_{\rm bias}$ and to the input of current mirror ($\rm M_2$–$\rm M_3$) which once again draws $I_{\rm bias} + I_{\rm ac}$. The copied current is passed to node C, where a current source of value $I_{\rm bias}$ is connected in parallel with a current mirror ($\rm M_5$–$\rm M_8$). The mirror naturally only mirrors the positive phases of $-I_{\rm ac}$. In other words, half-wave rectification of the current $-I_{\rm ac}$ is carried out. A similar operation is carried out at node D giving the half-wave rectified positive phases of $+I_{\rm ac}$. By summing the two half-wave rectified signals at node E, the full-wave rectified audio signal is thus recovered.

There are a number of device sizing issues associated with the full-wave rectifier and current limiter. On one hand, keeping devices small will save area and allow better rectification of small signals at higher frequencies since there is less charge stored in the channel, while on the other hand, if the devices are too small, this will lead to an unacceptably large mismatch. Mismatch can lead to a dc output in the absence of an ac signal, or alternatively, a minimum level below which nothing is detected. The devices that should be kept small, if small signals are to be adequately rectified at the higher frequencies, are $\rm M_5$–$\rm M_8$ and

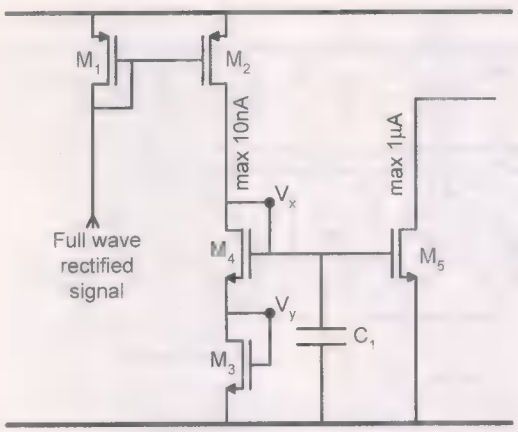


Fig. 13. This simple circuit performs low-pass filtering of the rectified signal, compression, and current amplification from nA to μA.

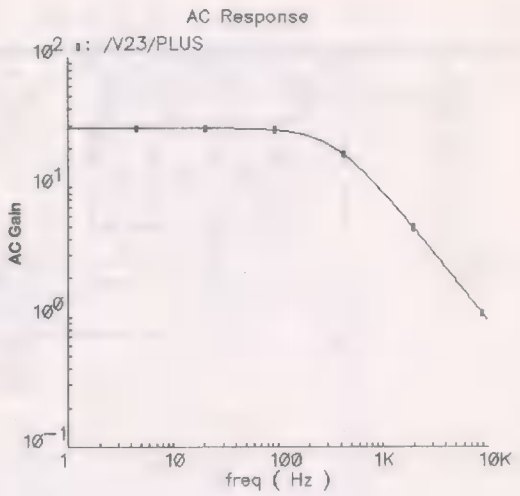


Fig. 14. Simulated AC response at 10 nA bias.

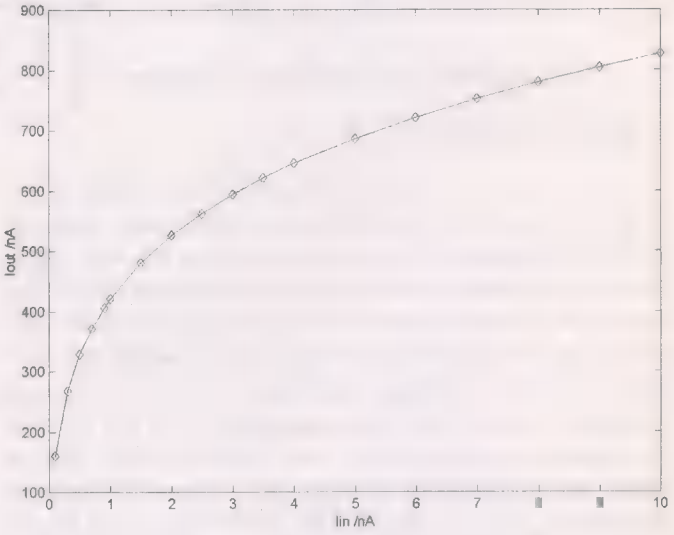


Fig. 15. Measured DC response at 10 nA bias.

M_9–M_{12}. Of course, the I_{bias} currents should be well matched as well as devices M_0–M_4. Fig. 12 illustrates the linearity of the rectifier over a couple of decades at 7 kHz. Performance improves at much lower frequencies than 7 kHz, which is where the filters are operating.

G. Low-Pass Filtering/Compression Amplification

In the conventional implementation of the CIS strategy, after the full-wave rectification comes the low-pass filter, which is followed by a separate signal compressor that reduces the dynamic range to fit the patient's low stimulation range. Classical techniques for low-pass filtering at the low frequencies of interest either consume much area due to large capacitors/resistors or consume more power when using active components to make small capacitors "appear large" via the Miller effect. The proposed solution is shown in Fig. 13.

If we exclude the current mirror (M_1–M_2) that is used for interfacing to the full-wave rectifier, with just three transistors and a capacitor, the full-wave rectified signal can be smoothed out, compressed, and amplified to stimulation levels. The maximum current input from the current-limited full-wave rectifier is 10 nA, which flows in all devices except for M_5, which boosts the signal to just under a microampere at maximum. Hence, this provides three signal-processing functions at a very low power budget. Inaccuracies due to process variations are not important since the patient-to-patient variations are much larger and are accommodated in the patient-fitting circuits that follow. In Fig. 14, the ac response at a bias of 10 nA illustrates a low cut-off frequency of just over 300 Hz with a 40–pF capacitor. The bias current supplied to this block *is* the full-wave rectified signal provided from the previous stage and so the cutoff frequency fluctuates to lower values accordingly. For example, at 500-pA current, the cutoff frequency crawls down to 40 Hz.

The measured dc response of the circuit is shown in Fig. 15, illustrating both dynamic-range reduction as well as current gain, taking the signal from the nanoampere range to microampere levels. Dynamic range compression means this gain is higher for smaller signals and lower for larger signals. By sizing the devices appropriately, during normal operation all transistors (10 μm$\times$ 10 μm) are in the subthreshold region except for M_5 (2 μm$\times$ 40 μm). Therefore, it is quite straightforward to derive an expression describing the circuit's dc input-output characteristic (neglecting the body effect):

$$I_{\text{out}} = \frac{\mu C_{ox} W_5}{2L_5}\left[2nV_{\text{T}} \ln\left(\frac{I_{in}}{\frac{W_{3,4}}{L_{3,4}} I'_o}\right) - V_{TH}\right]^2. \quad (4)$$

A number of different compression schemes are utilized in cochlear implant processors; it is not imperative that these are logarithmic, so it does not matter if the circuit of Fig. 13 does not perform a purely logarithmic compression. A more generalized form of compressions [23] used in cochlear implants is

$$I_{\text{out}} = Ax^P + K. \quad (5)$$

That concludes the last of the analog signal-processing functions. The compressed power in each frequency band is then sent to the patient-fitting and stimulation circuits.

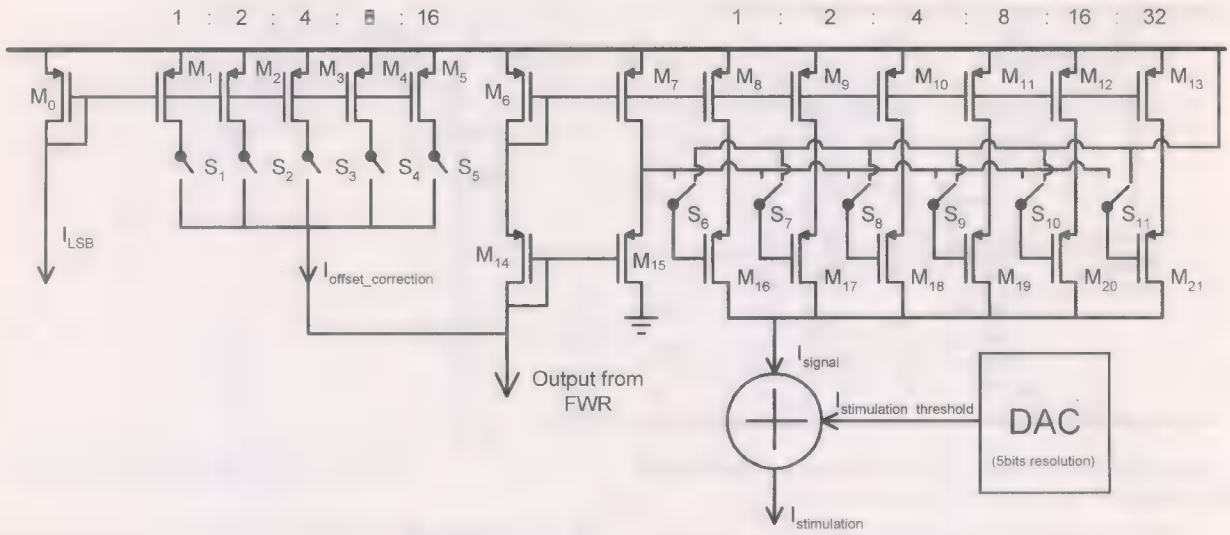


Fig. 16. Schematic of electrode driving circuits. Device M_6 has a smaller aspect ratio M_8. Devices M_7 and M_{15} are used to generate a bias voltage for the cascade devices M_{16}–M_{21} to increase output impedance without losing too much voltage headroom.

IV. Patient-Fitting and Stimulation Circuits

A. Patient-Fitting Circuits

The stimulation current levels required for a patient to just about perceive sound (stimulation threshold) varies quite significantly from patient to patient or even from electrode site to electrode site within the same cochlea. This greatly depends on the number of surviving neurons and the proximity of the electrodes to the nerves. Similarly, the maximum comfortable stimulation level also has a large variability. However, in all cases, the dynamic range between the hearing threshold and the maximum comfortable level is quite low, ranging typically from 6 to 20 dB. Hence, good fitting is important if the most is to be made of the limited dynamic range of the patient. The fitting circuits consist of digitally controlled variable-width current mirrors, as shown in Fig. 16. For each channel, 5 bits are allocated to removing any accumulated offsets from all the ASP circuits, since MOS devices biased in the subthreshold operating region have poor current-matching characteristics in comparison to identical devices operated in strong inversion [12]. Another 5 bits are allocated to setting the threshold of hearing, while another 6 bits are allocated for a multiplicative constant that takes the maximum allowable current level, leaving the full-wave rectifier and smoothing circuits, to the maximum comfortable stimulation level.

The same offset removal mechanism can also be used to reduce the sound window's capture dynamic range in noisy environments. In the highly successful n-of-m stimulation strategy, only the n strongest frequency bands, of a total of m separate channels, actually stimulate the cochlea. Once the offset is removed, the threshold of hearing is set via a second dc current source. Any ac power detected is added to this to give hearing sensation. Since the maximum current is limited at an earlier stage, the gain at the output is programmed such that at maximum input volume, the stimulus does not exceed the patients' comfortable hearing levels for each particular frequency.

B. CIS Biphasic Pulse Generation

The continuous interleaved sampling (CIS) generator is the last of the signal conditioning blocks that directly interfaces with the electrodes, via blocking capacitors. The CIS generator converts the output of the patient dynamic range mapping circuits into nonoverlapping biphasic pulses.

A top-level block diagram of the CIS generator is shown in Fig. 17. As there are 16 channels in the system, there are 16 output driver cells making up the CIS generator, however, only the first two and last two cells are shown. All the intermediate cells are identical, while the first and last cells differ slightly. The three different cells are shown in Fig. 18. Starting from the front cell, assuming there is no busy signal output from any of the following 15 cells or from within itself, the first cell will activate itself by generating a pulse with the three input NOR gate driving the first D-flip-flop input. A clock period later, the pulse will propagate to the output of the first flip-flop, which will turn on switches such as to provide a current path via M_1 to electrode A, back through electrode B, and down M_2 to ground. The first flip-flop is high so the three-input NOR gate will not produce another pulse as its output. After another clock cycle, the pulse will propagate on, flip-flop down, and reverse the direction of the current through the electrodes for another period. On the next clock pulse, a middle cell is activated and propagates the pulse in a similar fashion, first through itself and then down the remaining 13 cells, until it activates the last cell. This works in a similar fashion but has an extra flip-flop added to it so that it can provide an extra pulse that shorts all electrodes to ground, so as to remove any residual charge. This is required to make absolutely sure that no dc charge accumulates on the blocking capacitors, reducing voltage compliance. If blocking capacitors are not used and dc charge accumulates on the electrodes electrolysis may occur, corroding the electrodes and producing toxic materials, e.g., $2Cl^-$ ions could be turned into Cl_2 gas!

The clock used to drive the CIS generator is obtained from a simple *RC* relaxation oscillator shown in Fig. 19. This consists

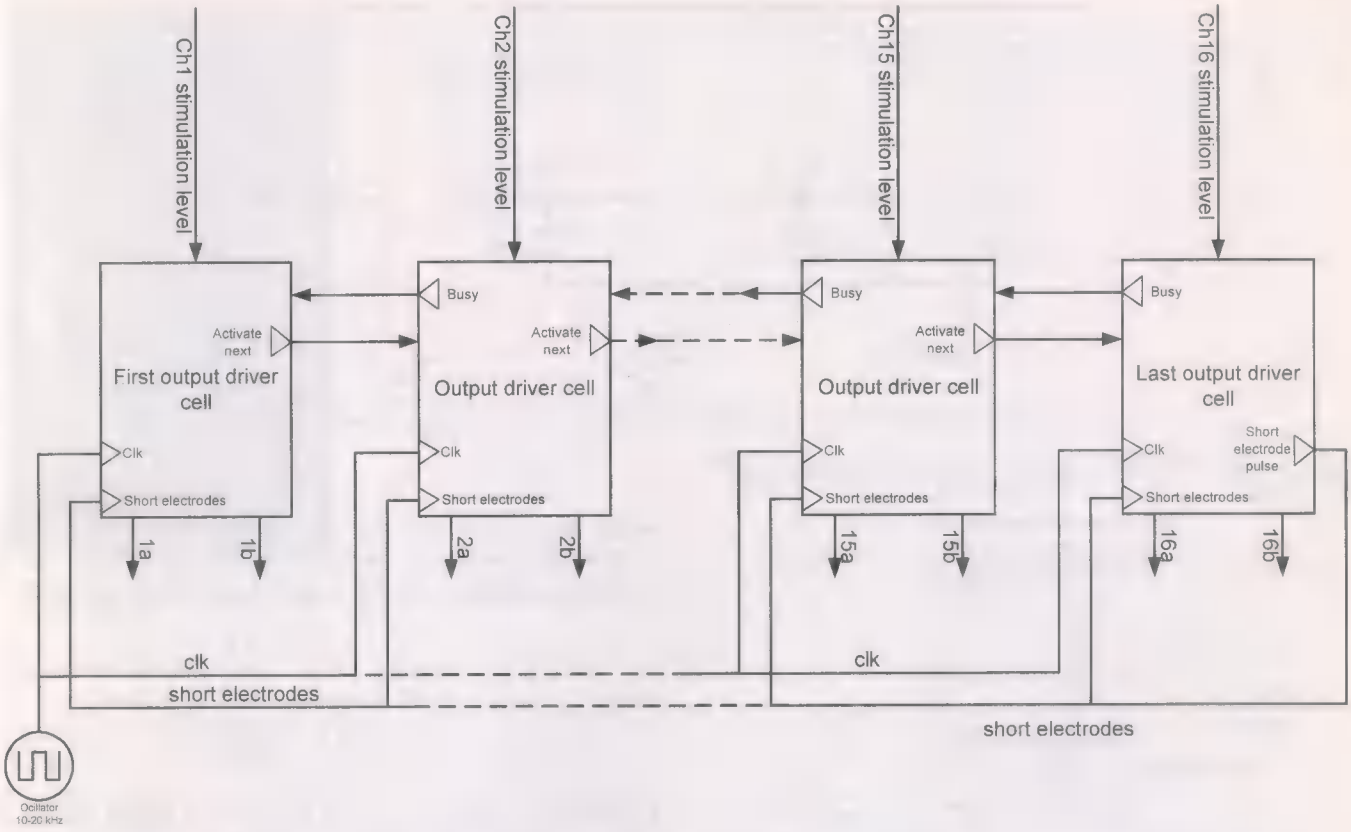


Fig. 17. Top-level view of the CIS generator circuit (only first two and last two channels).

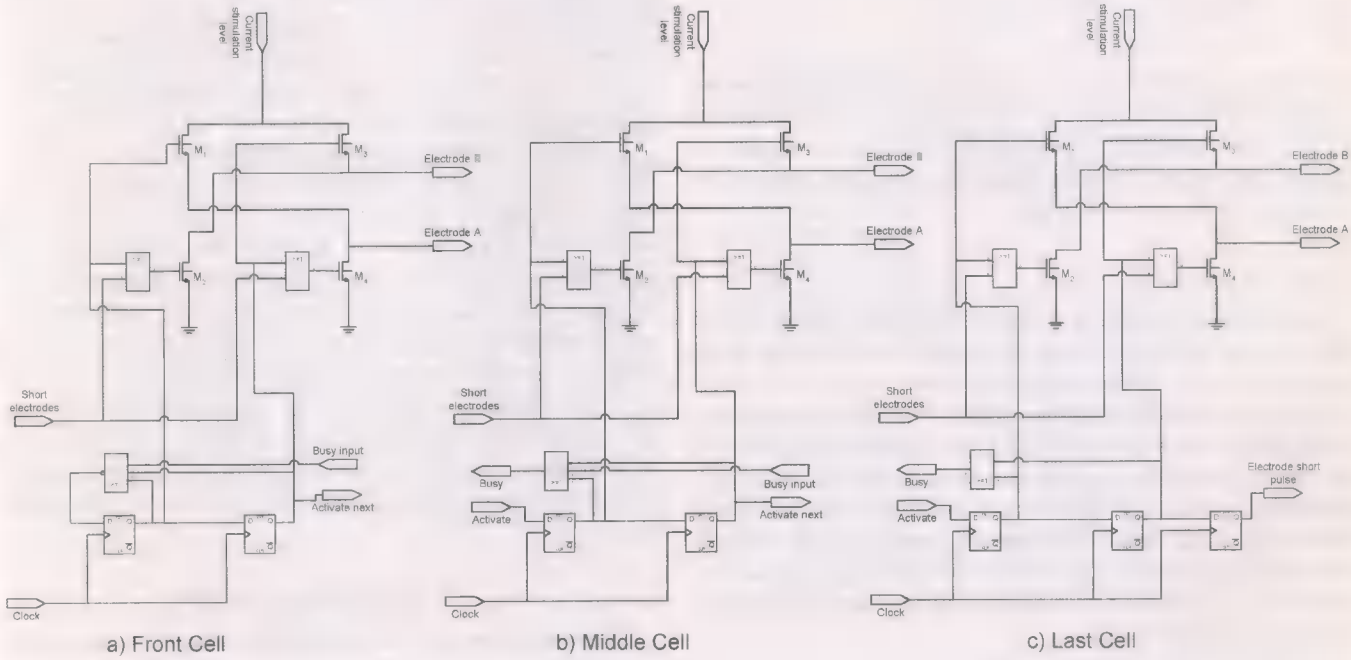


Fig. 18. The first, middle, and last cells making up the CIS stimulation generator are shown. By using a modular design, any number of channels can be easily assembled.

of three inverters, a capacitor, and a digitally controlled resistor that is used to adjust the frequency of oscillation. The frequency of oscillation directly affects the pulse width and the refresh rate of each channel.

V. Auxiliary Circuits

A. Power and Data Transfer

Power is sourced to the implant via an inductive link. The same inductive link is used to convey digital data to setup the

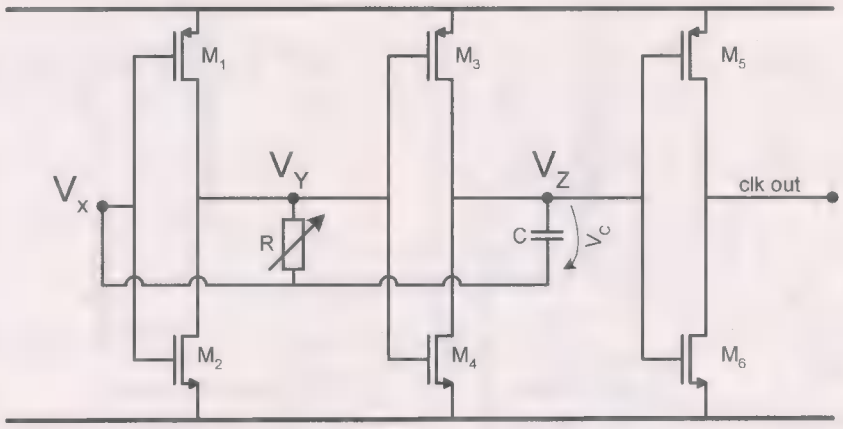


Fig. 19. Simple *RC* oscillator used to drive biphasic pulse generator.

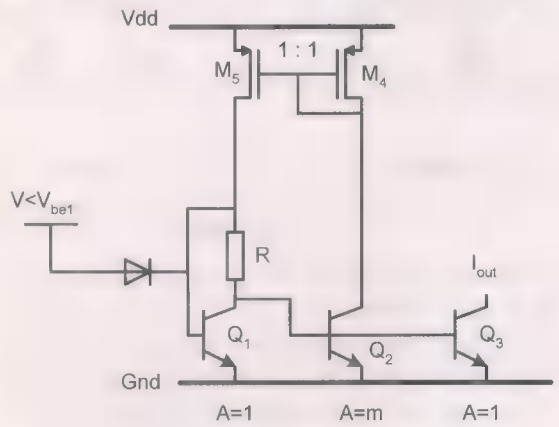


Fig. 20. Peaking current reference circuit.

system for a particular patients needs. The electromagnetic wave is pulse-width modulated in a similar fashion to that described in [24] which maintains a constant flow of power and data. A detailed description about the data recovery circuits and system setting can be found in [12], [25].

B. Reference Circuits

The reference circuits in a totally implanted system do not have to be particularly accurate but should remain consistent between the implant fitting/adjustment sessions so as not to overstimulate or understimulate the patients' neurons. Overstimulation results in exceeding the maximum comfortable level and can cause irreversible damage to surviving neurons, while understimulation does not make use of the already limited neural interface dynamic range. The implant is subjected to virtually no temperature variations, while the current design's power dissipation in of the order of microwatts and so does not affect the temperature within the casing.

The circuit chosen for the cochlear implant current reference is shown in Fig. 20. It is a low-current implementation of the peaking current reference source [26], [27]. The circuit achieves some degree of supply insensitivity by current feedback. Assuming $V_{\rm dd}$ rises, due to the finite output resistance of Q_2, $I_{\rm d4}$ rises too. This increase in current is mirrored and driven through resistor R. To accommodate the change in current, $V_{\rm beQ1}$ will increase logarithmically, while the voltage drop across the resistor will increase linearly. This causes $V_{\rm beQ2}$ to decrease, countering the initial increase in current, due to increase in supply voltage. It is quite simple to show that if

$$(W/L)_5 = (W/L)_4 \tag{6}$$

then

$$I_{\rm out} = \frac{kT}{eR}.\frac{1}{m}\ln m. \tag{7}$$

Letting $m = 8$ and R $= 243$ kΩ gives an output current of roughly 27 nAs. This simulates to about 30 nAs due to secondary effects (e.g., the vertical npn transistors have an Early voltage of about 30 V). The diode connected to Q1 is normally reverse biased except in startup conditions. The bias voltage for the startup diode was generated using a string of diodes connected between the supplies. Enough of them were used so as to ensure very little power loss. The current ranges from 270 pA to I nA when the supply varies from 3.8 to 4.2 V. The total power of the circuit including the outputs stands at 1.8 μW at 4 V.

The 1.1-V reference voltage required for the microphone supply is made by forward biasing a couple of diodes with around 10 nA of current to generate the voltage, and then buffering it. An off-chip capacitor is used to reduce the noise of this supply. $I_{\rm out}$ showed that it has 0.8% variation over supply voltage while $V_{\rm ref}$ has a 0.04% variation over the supply range. In terms of manufacturing variability, the circuit has an 11% standard deviation mainly due to resistor tolerances, but is digitally adjusted back to the nominal value. The digital adjustment is required in any case since inadequate electrode insertion requires the bias currents to be adjusted to compensate for this.

VI. The Overall System

The complete system fits on a die 3.5 mm $\times$ 6 mm; a photograph of the completed chip is shown in Fig. 21, along with a layout map. The top end of the chip contains the lowest power and noise components, while the noisiest and highest power circuit elements are placed on the bottom. A wide p^+ guard separates the predominantly digital circuitry from the analog circuitry. Six different supply pad pairs were used; for either half of the chip, a low-noise analog supply and separate digital supply was used. The fifth supply was used for the substrate

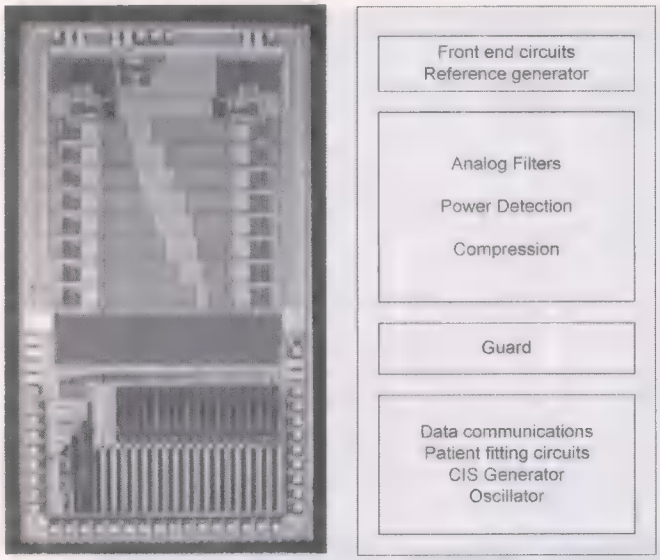


Fig. 21. Photograph of chip measuring 3.5 mm × 6 mm and layout plan.

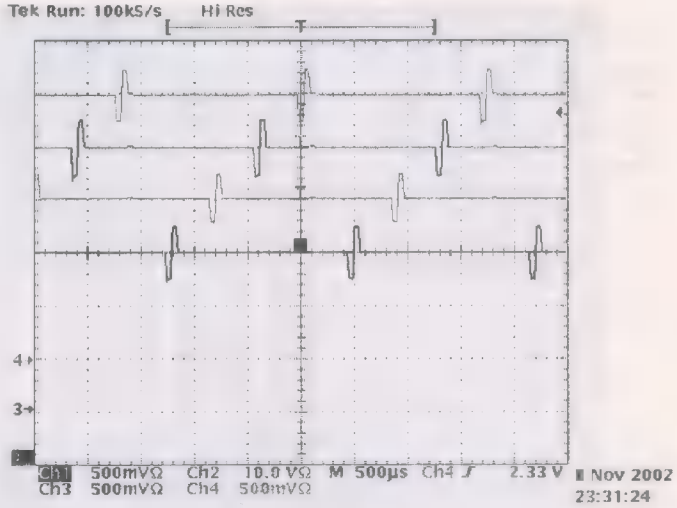


Fig. 22. Illustration of biphasic pulses output from channels 2, 6, 10, 14 of the whole system. The pulses were sent through a blocking capacitor and a resistor.

biasing/guard ring network of the low noise upper half of the chip; as the largest area is taken up by interleaved capacitors pairs, these were individually shielded from substrate noise by placing an n^- tub in the p-substrate beneath each one. This was connected to the positive bias supply. The last pair of supply pins provides power to the settings registry. When the battery supply is low, all the other supplies are cut off to prevent complete discharge. The static power consumption of the registry circuits is extremely low, in the order of femtoamperes. In the event that the registry power is completely cut off, on power up the system resets the registry's contents to ensure that the system comes up in a safe state, i.e., all outputs are set to zero.

In order to aid the testing of the cochlear system-on-chip, a dedicated test board was constructed. On the test board, a PIC microcontroller was used to send Hamming PWM encoded signals to control the settings on the chip. The biphasic current outputs of the stimulation circuits drove a resistor of similar impedance to that of a real cochlea, via a series blocking capacitor. The voltage developed across the resistor was amplified and sent to one of the PIC's A/D converters. As we can only monitor eight out of the 16 channels at any one time, the channels were split into odd and even channels with the use of dip switches on the test board. Fig. 23 shows the PC interface used to provide the settings on the cochlea chip. At the bottom, the resulting spectrogram is created by a log audio sweep, ranging in frequency from 10 Hz to 10 kHz. The intensity represents the magnitude of the current output pulses above the patients' threshold of hearing. The pulsating at the lower frequencies is due to the input signal frequency being comparable to that of the CIS output frequency. Table V contains a summary of the key features and performance characteristics.

The total power of the chip was measured to be 126 μW at 4 V, not including the power dissipated by the biphasic pulse stimulus. Assuming a battery capacity of 10 mA at 4 V on one charge, the circuit will be powered for about 13 days. Assuming that with the next generation electrodes that the stimulus is on average 500 μA in the constant presence of sound, then the total power will be 2.126 mW, so the same battery will last at least 18 hours and 48 mins which is quite reasonable, assuming that the implant is recharged during the patient's sleep.

TABLE V
FEATURES AND PERFORMANCE SUMMARY

General Characteristics

Process	AMS 0.8µm CXZ	
Die Area	3.5mm×6mm	
Number of channels	2×8 (logarithmically distributed)	
Power consumption	126µW (excl. electrode stimuli)	
AGC time constants (optional AGC and externally program. τ's)	τ_{attack} < 5ms, $\tau_{release}$ < 120ms,	
	Min	**Max**
Supply	3.8V	4.2V
Sound Pressure Level Range (noise free and clipping free)	30dB SPL	90dB SPL
Input dc voltage levels	0V	Vdd-500mV

Analog Signal Processing Characteristics

	Min	**Max**
Input Stage Dynamic Range	42dB	48dB
AGC Dynamic Range	14dB	30dB
Filter Dynamic Range	51dB	58dB
Filter Center Freq. Tuning	-20%	+50%
Smoothing Filter max f_C	-	300Hz
Compression Characteristic	$\propto (\ln)^2$	

Output Characteristics

Stimulation strategy	-Continuous Interleaved Sampling	
Stimulation pulse type	-Biphasic pulse ampl. Modulated	
Hearing threshold resol.	5bits (Max 100µA)	
Max comf. hearing resol.	6bits (Max 600µA)	
	Min	**Max**
Stimulation current	0	700µAs
Pulse width (Externally programmable to 4 levels)	50 µsec/phase	100µsec/phase

VII. DISCUSSION

The trend toward complete digital systems on chip has been re-examined to find that a hybrid digital analog system can save much more power for this particular application. In the above-

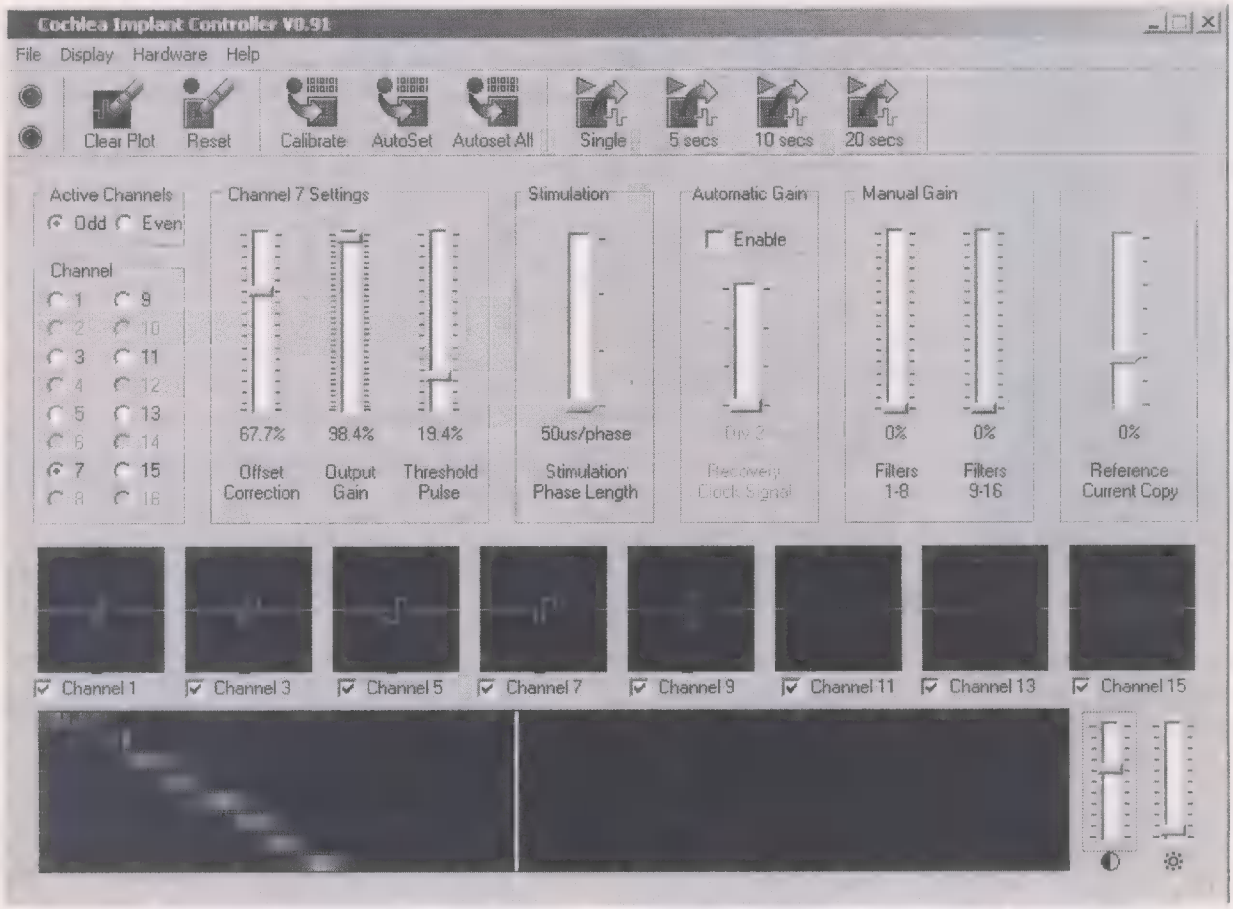


Fig. 23. Cochlear chip controller window interface. At the bottom is shown a spectrogram produced by a log frequency audio sweep between 10 Hz and 10 kHz.

described system, the power levels have been reduced from milliwatt levels that are currently used in cochlear chips to the microwatt range. A proof-of-concept design has been shown in solid-state form, however, before taking a system like this into production there is still much work to be done, e.g., in the areas of long-term reliability, patient safety through clinical trials, etc. Nevertheless, the design is based on existing successful cochlear implant processing strategies and so patient performance results are not expected to differ significantly.

REFERENCES

[1] F. Spelman, "The cochlear prosthesis: Review of design and evaluation of electrode implants for the profoundly deaf," *CRC Critical Views in Biomed. Eng.*, vol. 8, no. 3, pp. 223–252, 1982.

[2] E. A. Vittoz, "Low-power design: Way to approach the limits," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 1994, pp. 14–18.

[3] W. Germanovix and C. Toumazou, "Design of a micropower current-mode log-domain analog cochlear implant," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 47, no. 10, pp. 1023–1046, Oct. 2000.

[4] R. F. Lyon and C. A. Mead, "An analog electronic cochlea," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1119–1134, Jul. 1988.

[5] R. Sarpeshkar, R. F. Lyon, and C. Mead, "A low-power wide-dynamic-range analog VLSI cochlea," in *Analog Integrated Circuits and Signal Processing*. Boston, MA: Kluwer, 1996.

[6] E. Fragnière, "Analog emulation of the cochlea," Ph.D. dissertation, Swiss Federal Institute of Technology, Lausanne (EPFL), 1998.

[7] B. Wilson, D. Lawson, R. Wolford, and S. Brill, "Speech Processors for Auditory Prostheses," NIH, Fifth Quarterly Progress Report, NIH Project N01-DC-8-2105, 1999.

[8] A. Pavasovic, A. G. Andreou, and C. R. Westgate, "Characterization of subthreshold MOS mismatch in transistors for VLSI systems," *Analog Integrated Circuits and Signal Processing*, vol. 6, pp. 75–85, 1994.

[9] F. Serra-Graells, "VLSI CMOS subthreshold log companding analog circuit techniques for low voltage applications," Ph.D. dissertation, Instituto de Microelectrónica de Barcelona, Spain, 2001.

[10] J. Helms *et al.*, "Evaluation of performance with the COMBI 40 cochlear implant in adults: A multicentric clinical study," *Otorhinolaryngol.*, vol. 59, pp. 23–35, 1997.

[11] J. Kiefer *et al.*, "Speech understanding in quiet and noise with the CIS speech coding strategy (MED-EL COMBI 40) compared to the multipeak and spectral peak strategies (nucleus)," *Otorhinolaryngol.*, pp. 127–135, 1996.

[12] J. Georgiou, "Micropower electronics for neural prosthetics," Ph.D. dissertation, University of London, Imperial College of Science, Technology and Medicine, London, U.K., 2003.

[13] D. R. Frey, "Exponetial state space filters: A generic current mode design strategy," *IEEE Trans. Circuits Syst. I*, vol. 43, no. 1, pp. 34–42, Jan. 1996.

[14] C. Toumazou, J. Ngarmnil, and T. S. Lande, "Micropower log domain filter for electronic cochlea," *Electron. Lett.*, vol. 30, pp. 1839–1841, Oct. 1994.

[15] W. Germanovix, G. O'Neill, C. Toumazou, E. M. Drakakis, R. I. Kitney, and T. S. Lande, "Analogue micropowered log-domain tone controller for auditory prostheses," *Electron. Lett.*, vol. 34, no. 11, pp. 1051–1052, 1998.

[16] E. Seevinck, "Companding current-mode integrator: A new circuit principle for continuous-time monolithic filters," *Electron. Lett.*, vol. 26, pp. 2406–2407, Nov. 1990.

[17] E. M. Drakakis, A. J. Payne, and C. Toumazou, "Log-domain filters, translinear circuits and the Bernoulli cell," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1995, pp. 311–314.

[18] E. A. Vittoz, "Micropower techniques," in *Design of Analog-Digital VLSI Circuits for Telecommunications and Signal Processing*, J. E. Franca and Y. Tsividis, Eds. Upper Saddle River, NJ: Prentice-Hall, 1993.

[19] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low –voltage and low current applications," *Analog Integrated Circuits and Signal Processing*, vol. 8, pp. 83–114, Jul. 1995.

[20] R. M. Fox and M. Nagarajan, "Multiple operating points in log-domain filters," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 2, 1999, pp. 689–692.

[21] ——, "Multiple operating points in ▪ CMOS log-domain filter," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 6, pp. 705–710, Jun. 1999.

[22] J. Georgiou and C. Toumazou, "An operating point elimination technique for weak-inversion log-domain filters with multiple operating points," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 1, May 2001, pp. 153–155.

[23] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Mag.*, vol. 15, no. 5, pp. 101–130, Sep. 1998.

[24] J. A. De Lima, S. F. Silva, A. S. Cordeiro, A. C. Araujo, and M. Verleysen, "A low power silicon-on-insulator PWM discriminator for biomedical applications," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 5, Geneva, Switzerland, 2000, pp. 277–280.

[25] O. C. Omeni, "Efficient telemetry power/data links for biomedical applications," M.Sc. thesis, Imperial College of Science, Technology and Medicine, London, U.K., 2000.

[26] C. Y. Kwok, "Low voltage peaking complementary current generator," *IEEE J. Solid-State Circuits*, vol. 20, no. 3, pp. 8169–818, Jun. 1985.

[27] D. V. Kerns, "Optimization of the peaking current reference," *IEEE J. Solid-State Circuits*, vol. 21, no. 4, pp. 587–590, Aug. 1986.

Julius Georgiou (M'98) received the M.Eng. degree in electrical and electronic engineering in 1998 and the Ph.D. degree in electronics engineering from Imperial College of Science, Technology and Medicine, London, U.K., in 2003.

He was with the Micropower division of Toumaz Technology Ltd., Oxfordshire, U.K., as the Head of Design from 2001 to 2003, where he mainly worked on subthreshold circuit technology. Concurrently, he held an honorary position at Imperial College, U.K., where he lectured on a part-time basis. He is currently with Geosilicon Ltd, Nicosia, Cyprus.

Dr. Georgiou is a member of the IEEE Circuits and Systems Society, BioCAS Technical Committee, and the IEEE Circuits and Systems Society, Analog Signal Processing Technical Committee.

Christopher Toumazou (M'87–SM'99–F'01) received the Ph.D. degree from Oxford-Brookes University in collaboration with UMIST Manchester, U.K., in 1986.

He is a Professor of Circuit Design in the Department of Electrical and Electronic Engineering, Imperial College London, U.K. His research interests include high-frequency analog integrated circuit design in bipolar, CMOS, and SiGe technology for RF electronics, and low-power electronics for biomedical applications. He has authored or co-authored some 300 publications in the field of analog electronics and is a member of many professional committees. He is founder and Director of the Institute of Biomedical Engineering at Imperial, and the founder and Director of Toumaz Technology Limited.

Prof. Toumazou is a past Chairman for the Analog Signal Processing Committee of the IEEE Circuits and Systems (CAS) Society and past Vice-President of Technical Activities for the IEEE CAS Society. He is currently on the IEEE Society's Board of Governors. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–PART II: ANALOG AND DIGITAL SIGNAL PROCESSING, and is currently Honorary Editor and Chairman of the U.K. IEE *Electronics Letters*. He is co-winner of the IEE 1991 Rayleigh Best Book Award for *Analog IC Design: The Current-Mode Approach*. He is also a recipient of the 1992 IEEE CAS Outstanding Young Author Award for his work on high-speed GaAs op-amp design.

A 375 × 365 High-Speed 3-D Range-Finding Image Sensor Using Row-Parallel Search Architecture and Multisampling Technique

Yusuke Oike, *Student Member, IEEE*, Makoto Ikeda, *Member, IEEE*, and Kunihiro Asada, *Member, IEEE*

***Abstract*—A high-speed three-dimensional (3-D) image sensor for a 1000 range maps/s 3-D measurement system based on a light-section method is presented. It employs a row-parallel search architecture to achieve a high-speed frame access rate for the detection of activated pixels on the focal plane. The row-parallel search operation is carried out using chained search circuits embedded in a pixel. Moreover, we propose a row-parallel address acquisition technique using a bit-streamed column address flow. Row-parallel processors receive the bit-streamed column address and calculate the center position of activated pixels. The pipelined operations enable a multisampling technique that improves the resolution of pixel detection. A 375 × 365 3-D image sensor using the present architecture has been designed in ■ one-poly five-metal 0.18-μm standard CMOS process and successfully tested. It attains ■ frame access rate of 394.5 kHz with four samplings, which corresponds to 1052 range maps/s. The multisampling operation improves the sub-pixel resolution to around 0.2 pixels and achieves a range accuracy of less than 1.10 mm at a target distance of 600 mm.**

***Index Terms*—CMOS image sensor, high range accuracy, high speed, light-section method, multisampling method, range finder, row parallel architecture, 3-D image sensor.**

I. Introduction

A HIGH-SPEED and high-resolution three-dimensional (3-D) imaging system has a wide variety of applications including gesture recognition, depth-key object extraction, position adjustment, computer vision and security systems. In recent years, we have often seen 3-D computer graphics in movies and televisions and handled them interactively using personal computers and video game machines. Moreover, ultra-high-speed range finding provides the possibility of additional applications such as shape measurement of structural deformation and destruction, quick inspection of industrial components, observation of high-speed moving objects, and fast visual feedback systems in robot vision.

Some 3-D range-finding image sensors have been presented for 3-D imaging applications based on the stereo-matching method [1], [2], the time-of-flight method [3]–[7], and the light-section method [8]–[13]. The stereo-matching method provides a simple system configuration with two or more cameras. The stereo-matching processing, however, requires a huge computational effort with a high pixel resolution, and the range resolution and accuracy depend on target surface patterns. It is also difficult for the time-of-flight method to provide high range accuracy due to the limitations on the phase detection speed of a pulsed light. On the other hand, the light-section method is capable of high-accuracy range finding and it is most suitable for precision shape analysis. A typical configuration of light-section range finding is shown in Fig. 1(a). A sheet laser beam is projected and scanned on a target object. An image sensor detects the positions of the reflected beam on the sensor plane. 3-D range data are calculated by the beam projection angle α_p and the beam incidence angle α_i based on triangulation as shown in Fig. 1(b). The beam incidence angle can be acquired by the position of the incident beam on the sensor. Therefore many frames are necessary for a 3-D range image during the beam scanning. For example, a 1000 range maps/s 3-D measurement system with a practical pixel resolution requires over 100-kHz frame access rate. It is difficult for conventional image sensors to realize such a high-speed frame access. Fig. 2 plots the trend of range finding speed and pixel resolution in the state-of-the-art high-speed image sensors [14], [15] and light-section 3-D range finders [10]–[13]. It also shows examples of high-speed range finding applications. The conventional 3-D range finders have achieved 40–50 kHz frame access rate for real-time 3-D imaging. However, the target area of 1000 range maps/s requires around 400-kHz frame access rate. Therefore, we have presented a concept of a row-parallel search architecture on the focal plane and demonstrated the possibility of 1000 range maps/s range finding with a practical pixel resolution [16].

This paper presents a 3-D image sensor with 375 × 365 pixels for a 1000 range maps/s 3-D measurement system based on the light-section method, which was reported in part at the IEEE

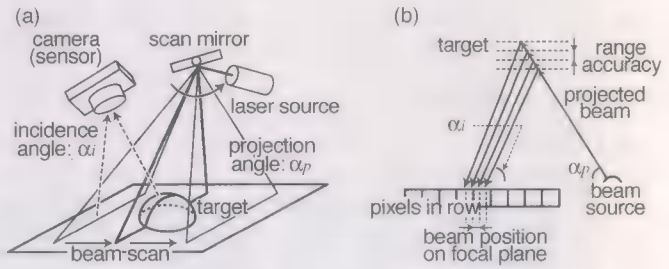


Fig. 1. Triangulation-based light-section range finding system. (a) System configuration. (b) Relation between range accuracy and beam position on the focal plane.

Manuscript received March 23, 2004; revised August 4, 2004.

The authors are with the University of Tokyo, Tokyo 113-8656, Japan (e-mail: y-oike@silicon.u-tokyo.ac.jp).

Digital Object Identifier 10.1109/JSSC.2004.841017

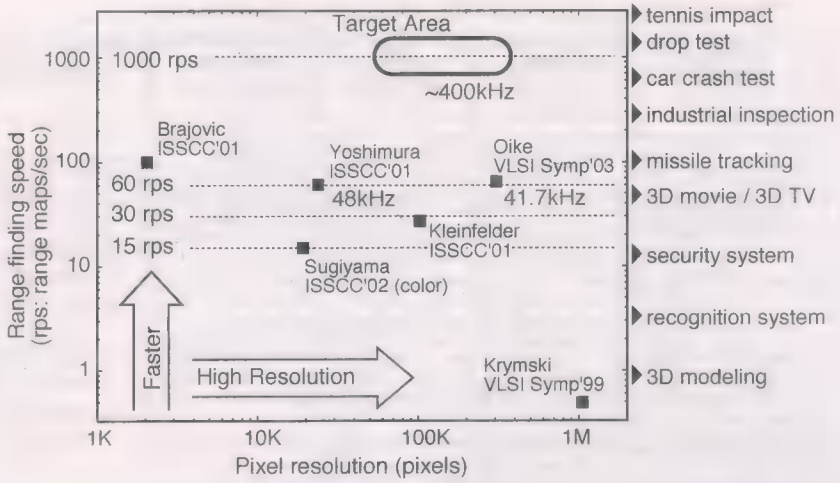


Fig. 2. Previous works and application examples.

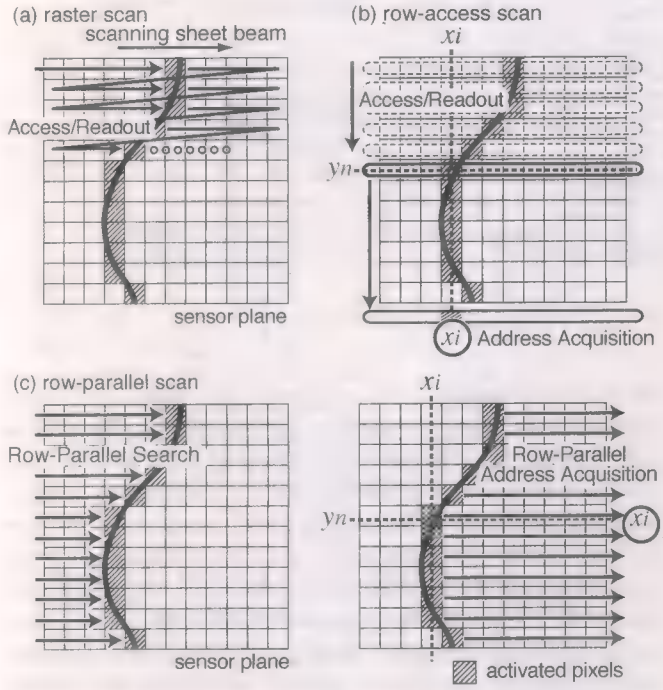


Fig. 3. Frame access methods. (a) Raster scan. (b) Row-access scan. (c) Row-parallel scan.

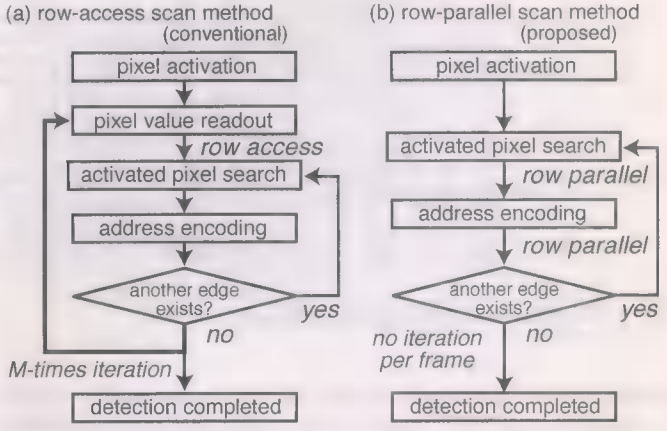


Fig. 4. Position detection flow. (a) Conventional row-access scan method with M row lines. (b) Proposed row-parallel scan method.

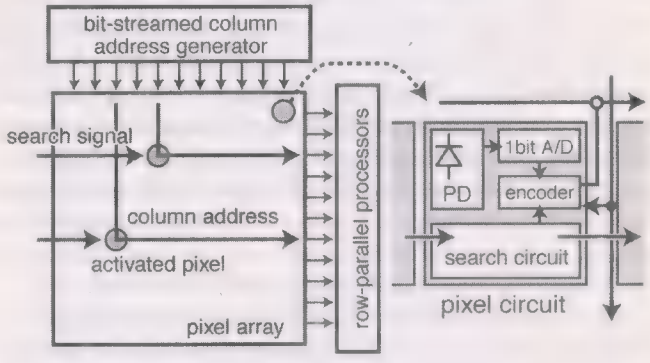


Fig. 5. Row-parallel position detection architecture implemented on the sensor plane.

ISSCC 2004 [17]. The row-parallel search architecture is implemented in three pipelined stages with a new multisampling function. The separated stages of photo integration, position detection, and data readout enable a high-speed frame access rate with multiple samplings. The multisampling technique improves the sub-pixel resolution of position detection on the focal plane for high range accuracy.

Section II presents the concept of a row-parallel search architecture. Circuit configurations and operations are described in Section III. Section IV introduces the multisampling technique with theoretical estimation of the improved sub-pixel resolution. Then, Section V shows the chip specification of a designed 3-D image sensor. The measurement results are discussed in Section VI. Finally, Section VII concludes this paper.

II. Row-Parallel Position Detection Architecture

A. Concept of Row-Parallel Search Architecture

Conventional image sensors typically employ a raster scan method or a row-access scan method. The raster scan method accesses all the pixels sequentially for a few activated pixels on the focal plane as shown in Fig. 3(a). The row-access scan method also needs to access all the pixel values. In row-access image sensors such as [11]–[13], the activated pixels in a row

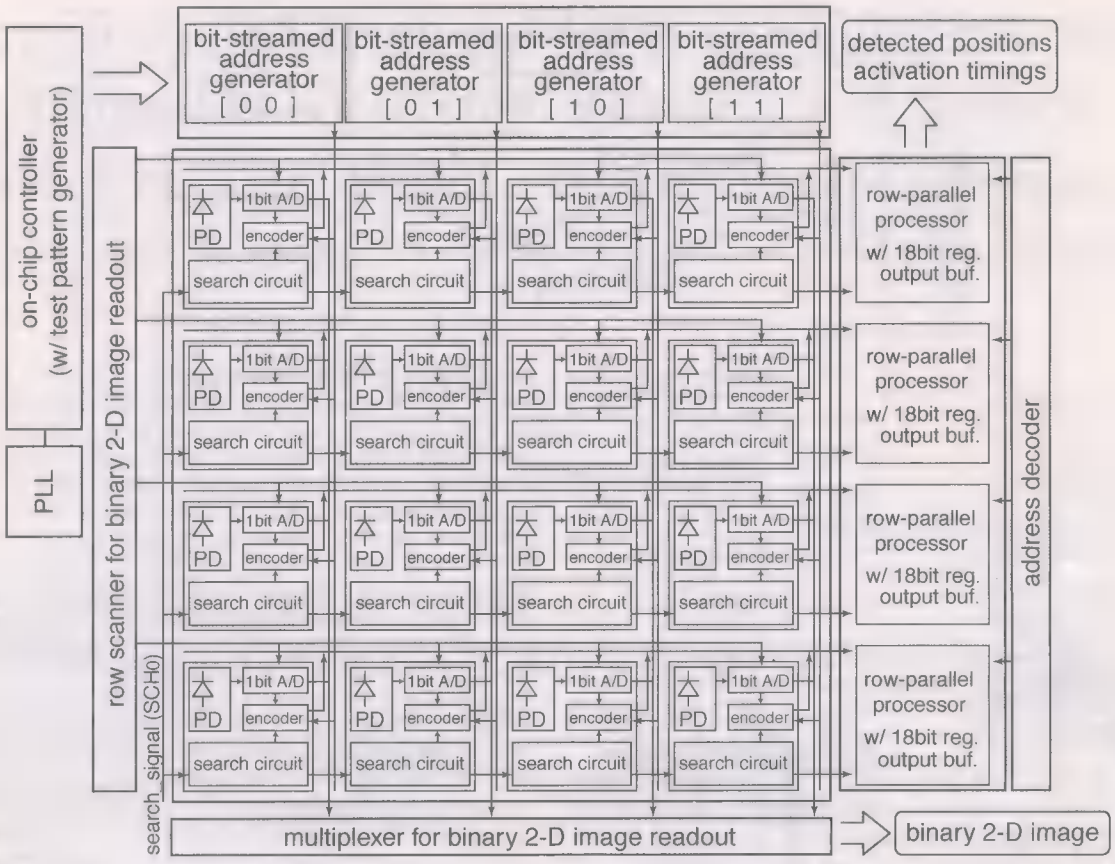


Fig. 6. Simplified block diagram of 4 × 4 pixels.

line can be scanned and detected in a column parallel fashion as shown in Fig. 3(b). Therefore, the row-access scan method is more suitable for high-speed position detection than the raster scan method. Fig. 4(a) shows the position detection flow of the row-access scan method. First some pixels are activated by a strong incident beam. Then the pixel values in a row line are read out. The activated pixels are scanned and detected in column parallel. The left and right edge addresses of consecutively activated pixels are acquired. If another incident beam exists in the row line, the search and address encoding operations are repeated. After that, the next row line is accessed and the pixel values are read out again. The access and search operations are repeated in proportion to the number of row lines. The access rate, limited to about 50 kHz, becomes the bottleneck.

Fig. 3(c) shows the proposed row-parallel scan method on the focal plane. In the row-parallel scan method, activated pixels in every row line are simultaneously scanned in row parallel. Then the addresses are acquired also in row parallel. Therefore there is no access iteration in proportion to the pixel resolution as shown in Fig. 4(b).

B. Block Diagram of Row-Parallel Scan Sensor

The present row-parallel architecture is implemented on the sensor plane as shown in Fig. 5. The row-parallel search operation is carried out by a chained search circuit embedded in each pixel. Search signals are provided from the left part of the sensor. They propagate from one pixel to the next pixel one after another via the in-pixel search circuit in a row parallel fashion. Then, the search propagation is interrupted at the first-encountered active pixel in each row line. In terms of address acquisition, it is impractical to implement an address encoder in every row line since a regularly spaced array structure is necessary for an image sensor. If a standard address encoder is implemented in each pixel, it requires many transverse wires per row as well as a large circuit area per pixel. We propose a bit-streamed column address flow for row-parallel address acquisition that enables compact circuit implementation. Column address streams are injected at the top part of the sensor in column parallel, and change their directions at pixels detected by the search circuits. The address acquisition scheme requires just one vertical wire per column and one transverse wire per row, which is suitable for a high-resolution pixel array. Each pixel includes a photo detector, a 1-bit A/D converter, a search circuit, and part of an address encoder.

Fig. 6 shows an overview of the row-parallel scan image sensor simplified to 4 × 4 pixels. It consists of a pixel array, bit-streamed column address generators at the top, row-parallel processors with data registers and output buffers on the right, a row scanner on the left, and a multiplexer at the bottom. These components are controlled by an on-chip sensor controller with a phase-locked loop (PLL) module. Pixels in a row line are connected with neighbor pixels by a search signal path. Column address streams are provided from the address generators to each vertical wire. Then the bit-streamed address signals are

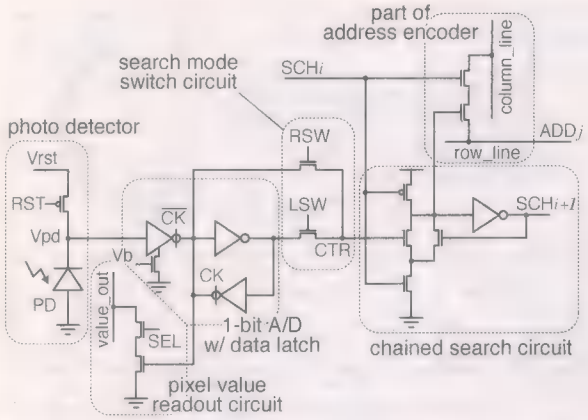


Fig. 7. Schematic of a pixel circuit.

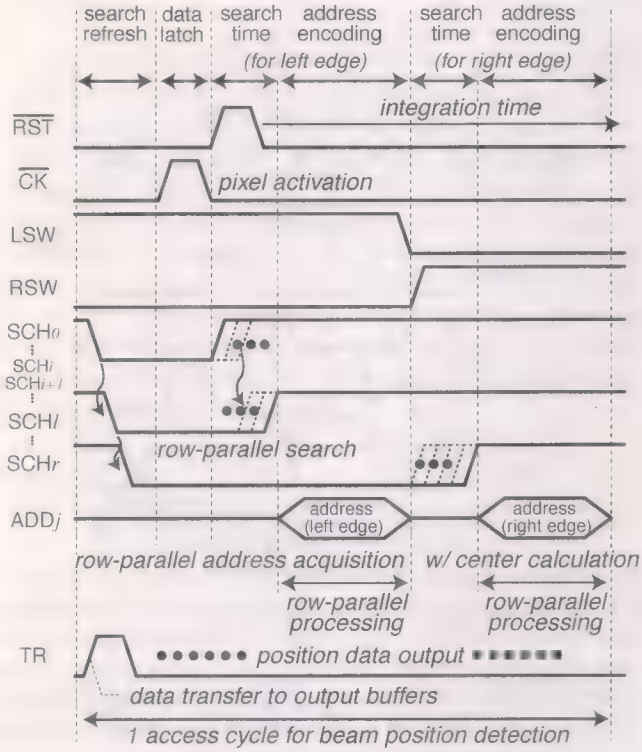


Fig. 8. Timing diagram of a pixel circuit.

injected to horizontal wires at the detected pixels. The row-parallel processors receive the bit-streamed address signals and the search completion signals from the right pixels in each row.

III. Circuit Configuration and Operation

A. Pixel Circuit Configuration

Fig. 7 shows the pixel circuit configuration with row-parallel position detection functions. It consists of a photo detector with a reset circuit, a 1-bit A/D converter with a latch circuit, a pixel value readout circuit, a search mode switch circuit, a chained search circuit, and part of an address encoder. The voltage V_{pd} is set to a reset voltage V_{rst} by *RST*. The 1-bit A/D converter receives V_{pd} and determines the pixel value. The voltage V_{pd} becomes a low level in case of an active pixel with strong incident intensity. Therefore, it provides "0" for an active pixel value, and

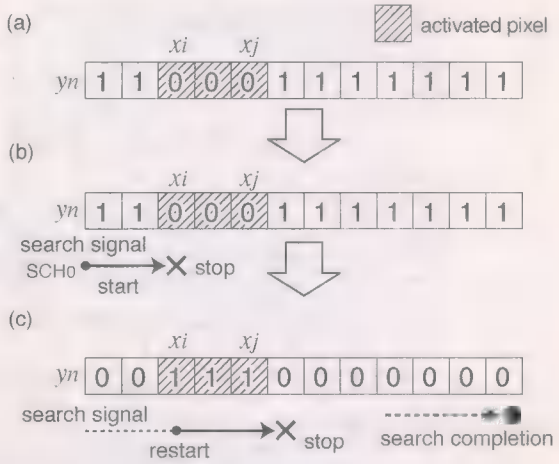


Fig. 9. Procedure of row-parallel activated pixel search. (a) Pixel activation; (b) search left edge; (c) invert all pixel values and search right edge.

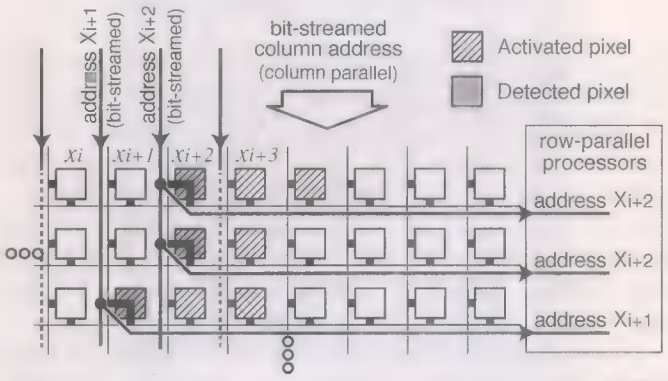


Fig. 10. Bit-streamed column address flow for row-parallel address acquisition.

"1" for an inactive pixel value. A transistor biased by V_b reduces the short-circuit current and controls the threshold level of A/D conversion. The pixel value readout circuit provides a binary image for functional tests. The search mode switch circuit and the chained search circuit are devoted to a row-parallel search for activated pixels. The address encoding section connects a column address line with a row address line. The row-parallel search and address acquisition functions are described in detail in the next sections.

B. Row-Parallel Search Operation

The row-parallel search operation is carried out using a chained search circuit embedded in each pixel. First, it detects the left edge of consecutively activated pixels in each row. Fig. 8 shows a timing diagram of the pixel circuit. Fig. 9 shows the procedure of the row-parallel search for activated pixels. The search mode switch circuit, which is implemented by a pass-transistor XOR, provides a control signal *CTR* for the search circuit. For the left edge detection, *LSW* and *RSW* are set to a high level and a low level, respectively. As the result of pixel activation, the activated pixel values are "0" and the others are "1"' as shown in Fig. 9(a). A search signal SCH_0 is provided to the left pixel in each row line. It passes through inactive pixels one after another via the in-pixel search circuits since the control signal *CTR* is set to a high level. The search

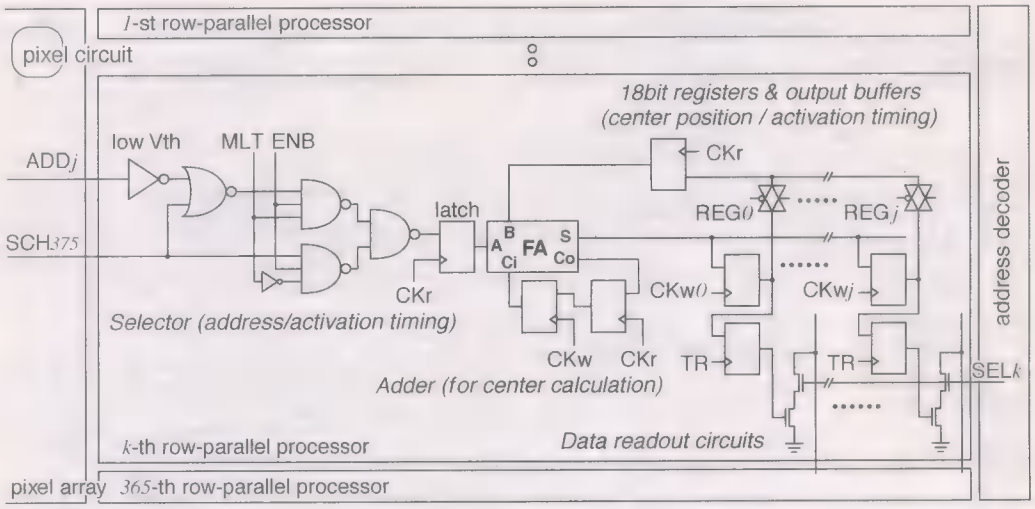


Fig. 11. Schematic of a row-parallel processor.

signal propagation is interrupted at the first-encountered active pixel as shown in Fig. 9(b), that is, it detects the left edge of consecutively activated pixels. After row-parallel address acquisition, *LSW* turns OFF and *RSW* turns ON. All the pixel values are inverted for the right edge detection as shown in Fig. 9(c). Namely, the active pixel values change to "1" and the interrupted search signal immediately starts again from the left edge. It passes through active pixels one after another and then stops at the next pixel of the right edge.

The worst delay of the search operation is the signal propagation delay through all the pixels in a row line. Therefore the search clock cycle is determined by the worst-case delay. The center position of incident beam can be calculated by the left and right edge addresses. The number of search cycles is the same regardless of the number of consecutively activated pixels. If another activated pixel exists on the same row, all the pixel values can be inverted again by switching *LSW* and *RSW*. The search operation restarts from the detected right edge to the next left edge. Therefore the row-parallel search operation is capable of position detection for multiple incident beams due to the search continuation. The last search signal SCH_n from the right pixel indicates whether no activated pixel exists in each row as a search completion signal.

C. Row-Parallel Address Acquisition

Fig. 10 shows a bit-streamed column address flow for row-parallel address acquisition. A column address line is connected to a row address line by part of an address encoder in the detected pixel. The row-parallel address acquisition needs just 2 pass transistors in a pixel as shown in Fig. 7. At the detected left edge, SCH_l from the previous pixel becomes a high level, but the next search signal SCH_{l+1} is still a low level since the search signal propagation is interrupted. Therefore, both inputs, SCH_i and $\overline{SCH_{i+1}}$, are set to a high level at the detected pixel. A bit-streamed address signal is then provided from a column address line to a row address line via the two pass transistors. The column address streams never conflict with each other in the same row line since the left or right edge is detected by the

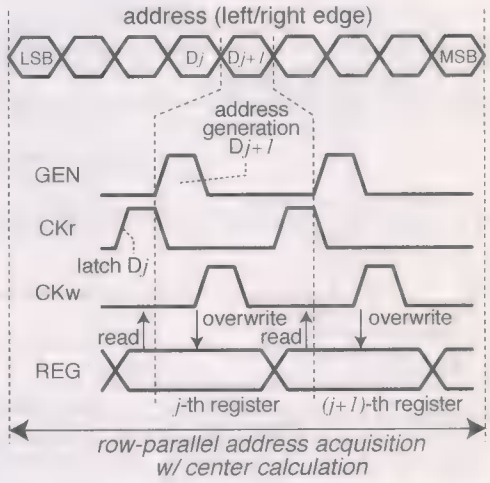


Fig. 12. Timing diagram of a row-parallel processor.

row-parallel search in each row. The bit-streamed address signals are injected from the LSB to the MSB, and then they are received by the row-parallel processors.

D. Row-Parallel Processing

The range-finding image sensor has row-parallel processors that receive bit-streamed address signals ADD_j and search completion signals SCH_{375} in each row. Fig. 11 shows a schematic of the row-parallel processor. It consists of a selector with a signal receiver, a full adder, 18-bit registers, 18-bit output buffers, and data readout circuits. The selector switches the processing functions, which are an address acquisition mode and an activation counting mode. Fig. 12 shows a timing diagram of the row-parallel processor. A bit-streamed address signal is received by a low-threshold inverter because the address signal cannot swing to the supply voltage due to pass transistors in a pixel. In a multisampling operation, the row-parallel processor counts the number of usable pixel activations by the search completion signal since an occasional search operation includes no activated pixel. The address acquisition mode and the activation counting mode are switched by *MLT*. The left edge

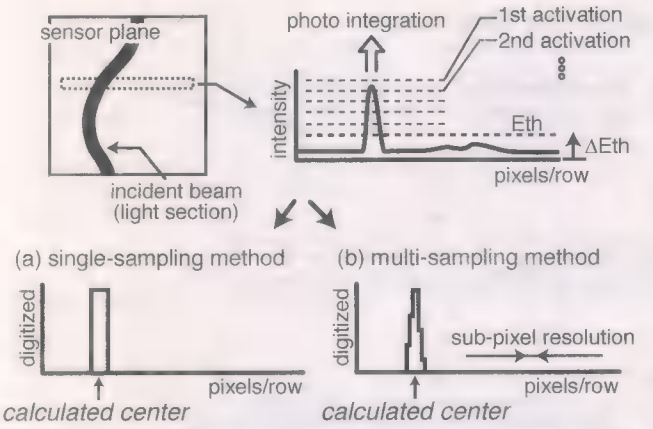


Fig. 13. Sub-pixel center position detection by multisampling method. (a) Single-sampling method. (b) Multisampling method.

address is stored in the registers. Then the right edge address is accumulated on the left edge address by CK_r and CK_w in sequential order from the LSB to the MSB. *ENB* is employed to disable the input of the full adder for carry accumulation in a multisampling operation. The accumulated address represents the center position of activated pixels. The results are transferred to the output buffers by *TR*, and then they are read out by SEL_k during the search operations for the next frame. The row-parallel processing is executed concurrently with the row-parallel address acquisition. The row-parallel processor has the capability to perform a multisampling operation due to the high-speed position detection.

IV. Multisampling Position Detection

Three-dimensional range data is calculated by the beam projection angle α_p and the incident angle α_i as shown in Fig. 1(b). The incident beam angle α_i is provided from the incident beam position on the focal plane. Therefore, the range resolution and accuracy depend on the resolution of position detection on the sensor. In other words, the sub-pixel resolution efficiently improves the range accuracy. A multisampling technique is implemented to acquire the intensity profile of incident beam for a fine sub-pixel resolution.

In a multisampling method, all the pixel values are updated repeatedly during the photo integration. Pixels with stronger incident intensity are activated faster and found many times in multiple samplings as shown in Fig. 13. In the conventional single sampling mode, the acquired data are binary, and so the sub-pixel resolution of calculated center position is 0.5 pixels as shown in Fig. 13(a). On the other hand, the number of samplings represents the scale of the intensity profile as shown in Fig. 13(b). Some scales provide a fine sub-pixel resolution of center position detection for range accuracy improvement. Fig. 14 shows a theoretical estimation of the sub-pixel resolution as a function of the number of samplings. A gaussian distribution is assumed as the beam intensity profile. The sub-pixel resolution is efficiently improved in 2–8 samplings. For example, a 4-sampling mode attains 0.2 sub-pixel resolution.

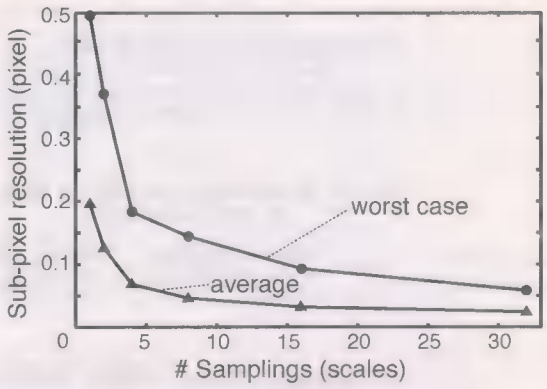


Fig. 14. Sub-pixel resolution as a function of the number of samplings.

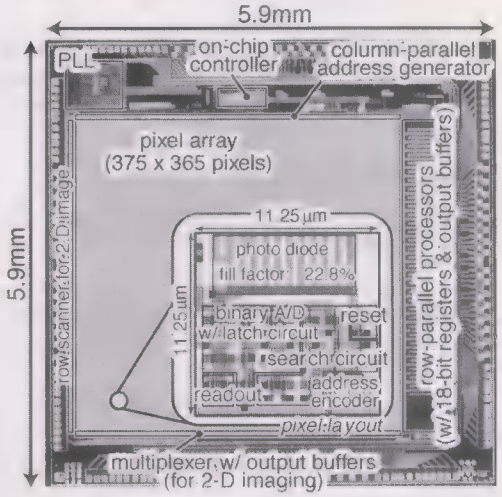


Fig. 15. Die microphotograph and pixel layout.

TABLE I
Chip Specifications

Process	1P5M 0.18 μm CMOS process
Die size	5.9 mm × 5.9 mm
Resolution	375 × 365 pixels
Pixel size	11.25 μm × 11.25 μm
Fill factor	22.8 %
Pixel configuration	1 PN-junction PD, 24 FETs / pixel
Total FETs	3.74 M transistors

V. Chip Implementation

A 375 × 365 3-D range-finding image sensor using the present row-parallel architecture has been designed and fabricated in a 0.18 μm standard CMOS process with 1-poly-Si 5-metal layers. The die size is 5.9 mm × 5.9 mm. Fig. 15 shows a chip microphotograph and a pixel layout. The sensor consists of a 375 × 365 pixel array, a column-parallel address generator, and row-parallel processors with 18-bit registers and output buffers. A row scanner and a column multiplexer are also implemented to acquire a binary 2-D image for test. The row-parallel operations are executed by an on-chip sensor controller with a PLL module. The implementation requires 3.74 million transistors. The supply voltage is 1.8 V. The pixel size is 11.25 μm × 11.25 μm with 22.8% fill factor. It consists of a PN-junction photo diode and 24 transistors. The photo diode is composed of n^+-diffusion and p-substrate. It is split into

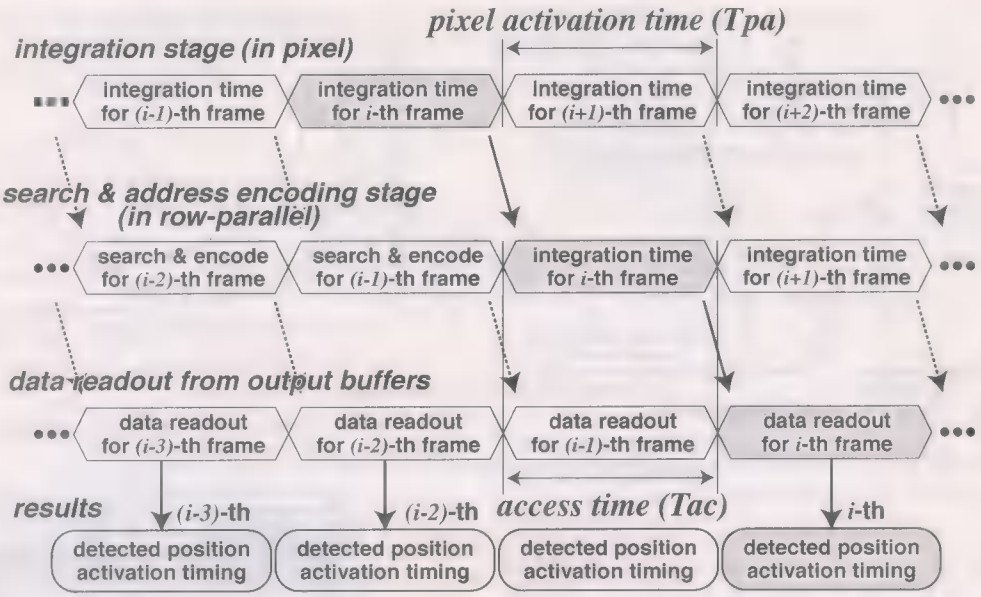


Fig. 16. Pipelined operation diagram.

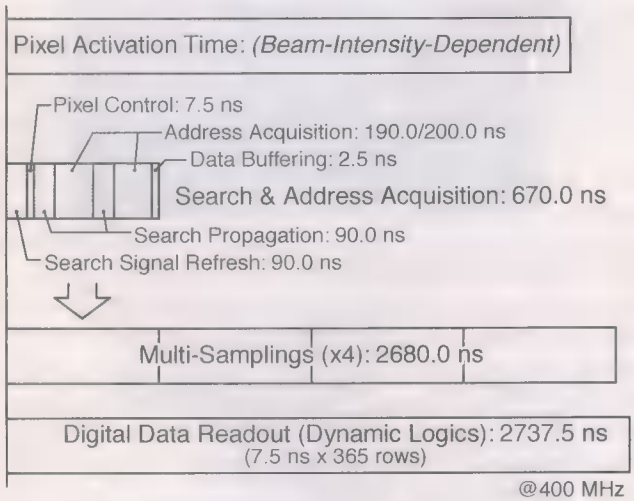


Fig. 17. Cycle time of activated pixel search and data readout.

several rectangular slices to improve the sensitivity since the present CMOS process has no option of silicide layer removal. Table I shows the chip specifications.

VI. Measurement Results

A. Frame Access Rate

The row-parallel position detection is pipelined in three stages on the sensor as shown in Fig. 16. The first stage is the photocurrent integration for pixel activation. The second stage is the row-parallel operation of activated pixel search and address acquisition. The last stage is the data readout operation from output buffers. The photocurrent integration period is called the pixel activation time. It depends on the incident beam intensity and the sensitivity of a photo diode. That is, the pixel activation time can be controlled by the beam intensity. On the other hand, the access time is limited by a search operation with address acquisition or a data readout operation. Therefore our principal aim is to achieve a short access time for high-speed position detection.

Fig. 17 shows a cycle time of each pipelined stage at a 400-MHz operation. The worst case of search signal propagation takes 90 ns. So the search path refresh and the search operations for the left and right edges each require 90 ns. The row-parallel address acquisition takes less than 200 ns in the worst case. The worst case of address acquisition occurs when all the detected pixels are placed on the same column because the load capacitance of a column address generator becomes largest and limits the injection speed of the bit-streamed column address signals. The total cycle time of search and address acquisition is 670 ns. The limiting factor of the access time is the digital readout stage from output buffers, which requires 2737.5 ns. Therefore, the search and address acquisition can be repeated four times in the data readout period while maintaining the frame access rate.

We have tested the maximum access rate of the designed sensor. The sensor allows user-specified pixel activation. The worst-case situation is set by an electrical pattern on the sensor plane. Fig. 18 shows measured waveforms of the worst-case frame access to an electrical test pattern at 432 MHz. Fig. 19 shows a data readout circuit and the test equipment that was used for probing the output signals. Output buffers in each row are selected by SEL_k. The position results are read out by the dynamic readout circuits where are precharged by *PRE*, and received by sense amplifiers that are synchronized with *SACK*. The reference voltage $V_{\rm ref}$ is set to 300 mV below the supply voltage. The output signals are probed with parasitic capacitances of C_{IN} and C_{PB}, which are 7 and 13 pF, respectively. All the activated pixels are set in the 374-th column as the worst-case situation. The expected results were successfully acquired up to 432-MHz operation. The image sensor attains a frame access rate of 394.5 kHz, which corresponds to 1052 range maps/s

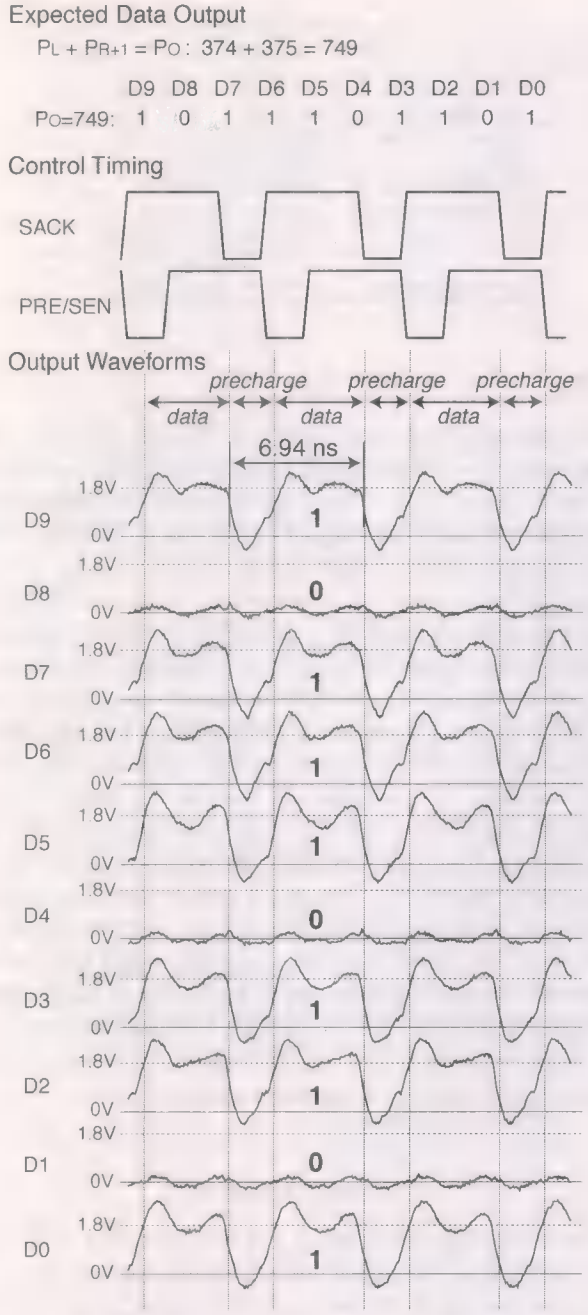


Fig. 18. Measured waveforms of the worst-case frame access to electrical test pattern at 432 MHz.

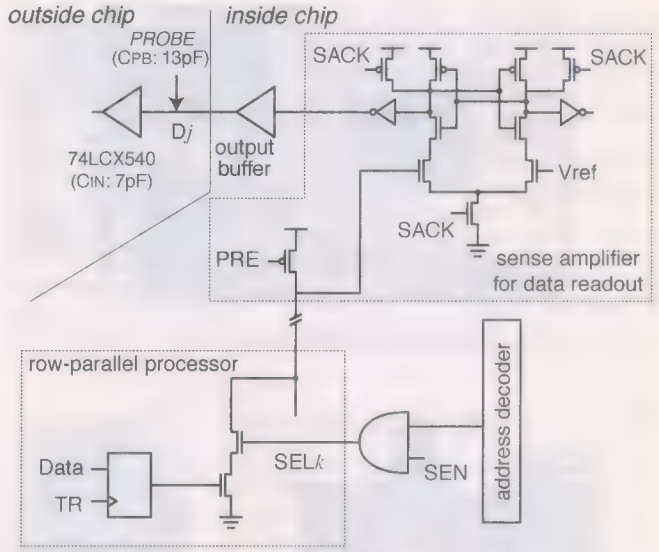


Fig. 19. Test equipment for the worst-case frame access.

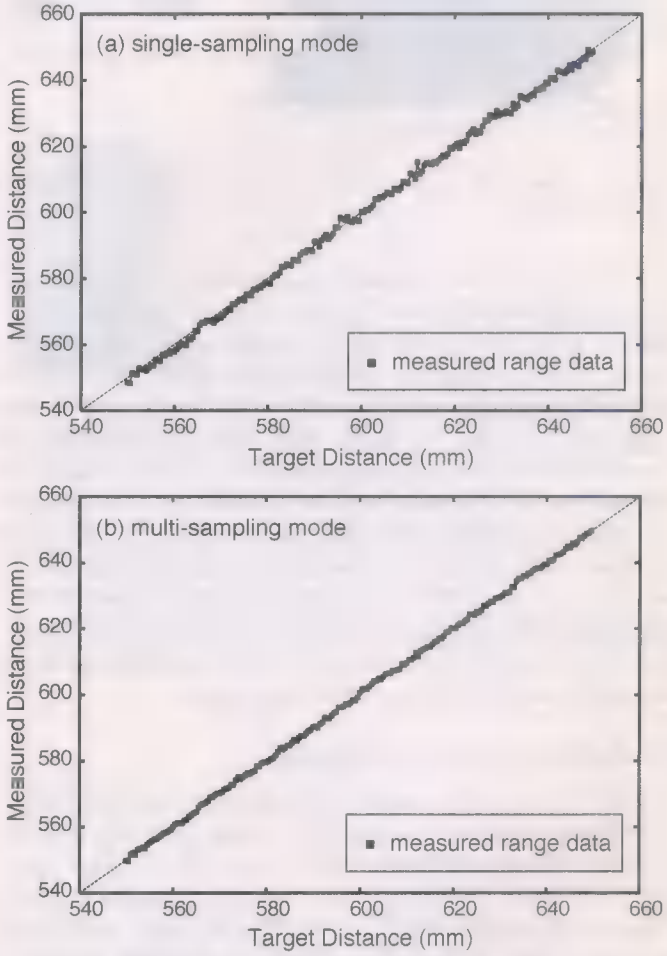


Fig. 20. Measured range accuracy. (a) Single-sampling mode. (b) Multisampling mode.

with 375 × 365 range data. The data rate is 144 Mbit/pin·s in the maximum frame access rate.

B. Range Accuracy

Fig. 20 shows the measured range accuracy at a target distance of around 600 mm. The *X* axis represents target distance and the *Y* axis represents measured distance. Fig. 20(a) shows the measured results in the conventional single sampling mode. The maximum range error is 2.78 mm and the standard deviation of error is 1.02 mm. The conventional single sampling mode achieves 0.46% range accuracy with 0.5 sub-pixel resolution. The range error is typically dominated by the pixel quantization error of position detection on the focal plane. Therefore, the range error can be suppressed by the multisampling technique with four scales as shown in Fig. 20(b). The maximum range error is 1.10 mm and the standard deviation is 0.47 mm in the same situation. The multisampling mode attains 0.18% range accuracy, which corresponds to around 0.2 sub-pixel resolution.

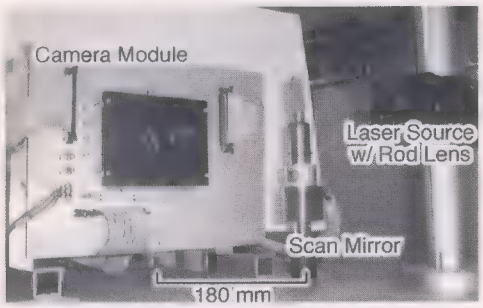


Fig. 21. Photograph of a range finding system.

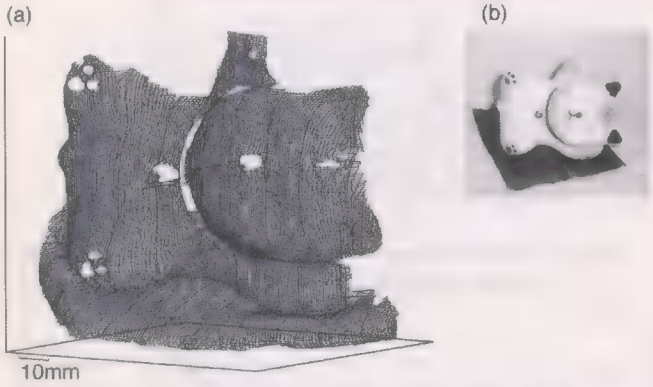


Fig. 22. Measurement result of range finding. (a) Measured range data. (b) Target object.

The range accuracy suffers from fluctuation of the threshold voltage of pixel activation. The peak-to-peak threshold fluctuation is about 150 mV including the reset voltage drop on the sensor, which is calculated by binary 2-D images that are measured using various reset voltages. However, the intensity profile with four scales does not fatally suffer from the fluctuation because the fluctuation has strong correlation with the location on the sensor and it is small enough to still allow the calculation of the center position in a local area. The timing of pixel activation is separated from the search and address acquisition operations as shown in Fig. 8. That is, the pixel activation is executed after the search path refresh and before the search signal propagation. Therefore, the pixel activation is not affected by crosstalk caused by digital signaling on the focal plane.

C. Example of Measured Range Image

Fig. 21 shows a photograph of the present measurement setup. The baseline between a camera and a beam projector is set to 180 mm. The target distance is 600 mm and the target scene is 90×90 mm^2. A 300-mW laser beam is expanded by a rod lens as a sheet beam with 5 mm width. The beam wavelength is 635 nm. Fig. 22 shows an example of measured range images. The measured 3-D data are plotted on three-dimensional coordinates as a wire-frame model (a) of a target object (b) in Fig. 22. In the present measurement setup, the limiting factor of the range finding is the pixel activation time. So the system requires a higher sensitivity photo detector or a sharp and strong laser beam. Our future work is to get better performance of the designed image sensor by satisfying these system requirements. Table II summarizes the chip performances.

TABLE II
CHIP PERFORMANCE

Supply voltage	1.8 V
Max. clock freq.	432 MHz
Frame access rate	394.5 kHz
Data rate	144 M bit/pin/sec
Range finding speed	1052 range maps/sec
Sub-pixel resolution	0.2 pixels (4 samplings)
Range accuracy	max. 1.10 mm @ 600 mm S.D. 0.47 mm @ 600 mm
Power dissipation	1065 mW @ 432 MHz, 1.8 V

VII. CONCLUSION

We have presented a high-speed 3-D image sensor for a 1000 range maps/s 3-D measurement system which has many potential applications such as shape measurement of structural deformation and destruction, quick inspection of industrial components, observation of high-speed moving objects, and fast visual feedback systems in robot vision. A row-parallel frame access architecture has been proposed for the high-speed range finding. The row-parallel search operations are executed by a chained search circuit embedded in a pixel on the focal plane. The bit-streamed column address flow enables row-parallel address acquisition with a compact circuit implementation. Moreover a multisampling technique is available for range accuracy improvement. A 375×365 3-D range-finding image sensor has been designed and fabricated in a one-poly five-metal (1P5M) 0.18-μm standard CMOS process. It attains a high-speed frame access rate with multiple samplings. The maximum frame access rate is 394.5 kHz with four samplings, which has a potential capability of 1052 range maps/s in the case of a sufficiently strong beam intensity. Then it provides 1.10 mm range accuracy at a target distance of 600 mm. It has been improved up to 0.2 sub-pixel resolution by the multisampling technique.

ACKNOWLEDGMENT

The VLSI chip in this study has been designed with CAD tools of Synopsys Inc. and Cadence Design Systems Inc., and fabricated through VLSI Design and Education Center (VDEC), University of Tokyo, in collaboration with Rohm Corporation and Toppan Printing Corporation.

REFERENCES

[1] T. Kato, S. Kawahito, K. Kobayashi, H. Sasaki, T. Eki, and T. Hisanaga, "A binocular CMOS range image sensor with bit-serial block-parallel interface using cyclic pipelined ADC's," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2002, pp. 270–271.

[2] R. M. Philipp and R. Etienne-Cummings, "Single chip stereo imager," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, May 2003, pp. 808–811.

[3] R. Miyagawa and T. Kanade, "CCD-based range-finding sensor," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1648–1652, Oct. 1997.

[4] P. Gulden, M. Vossiek, P. Heide, and R. Schwarte, "Novel opportunities for optical level gauging and 3-D-imaging with the photoelectronic mixing device," *IEEE Trans. Instrum. Measure.*, vol. 51, no. 4, pp. 679–684, Aug. 2002.

[5] R. Jeremias, W. Brockherde, G. Doemens, B. Hosticka, L. Listl, and P. Mengel, "A CMOS photosensor array for 3D imaging using pulsed laser," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2001, pp. 252–253.

[6] A. Ullrich, N. Studnicka, and J. Riegl, "Long-range high-performance time-of-flight-based 3D imaging sensors," in *Proc. IEEE Int. Symp. 3D Data Processing Visualization and Transmission*, Jun. 2002, pp. 852–856.

[7] M. Kawakita, T. Kurita, H. Hiroshi, and S. Inoue, "HDTV axi-vision camera," in *Proc. International Broadcasting Convention (IBC)*, Sep. 2003, pp. 397–404.

[8] A. Gruss, L. R. Carley, and T. Kanade, "Integrated sensor and range-finding analog signal processor," *IEEE J. Solid-State Circuits*, vol. 26, no. 3, pp. 184–191, Mar. 1991.

[9] M. de Bakker, P. W. Verbeek, E. Nieuwkoop, and G. K. Steenvoorden, "A smart range image sensor," in *Proc. Eur. Solid-State Circuits Conf.*, Sep. 1998, pp. 208–211.

[10] V. Brajovic, K. Mori, and N. Jankovic, "100 frames/s CMOS range image sensor," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2001, pp. 256–257.

[11] S. Yoshimura, T. Sugiyama, K. Yonemoto, and K. Ueda, "A 48 k frame/s CMOS image sensor for real-time 3-D sensing and motion detection," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2001, pp. 94–95.

[12] T. Sugiyama, S. Yoshimura, R. Suzuki, and H. Sumi, "A 1/4-inch QVGA color imaging and 3-D sensing CMOS sensor with analog frame memory," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2002, pp. 434–435.

[13] Y. Oike, M. Ikeda, and K. Asada, "640 × 480 real-time range finder using high-speed readout scheme and column-parallel position detector," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2003, pp. 153–156.

[14] A. Krymski, D. Van Blerkom, A. Andersson, N. Bock, B. Mansoorian, and E. R. Fossum, "A high speed 500 frames/s 1024 × 1024 CMOS active pixel sensor," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 1999, pp. 137–138.

[15] S. Kleinfelder, S. H. Lim, X. Liu, and A. E. Gamal, "A 10000 frames/s CMOS digital pixel sensor," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 2049–2059, Dec. 2001.

[16] Y. Oike, M. Ikeda, and K. Asada, "High-speed position detector using new row-parallel architecture for fast collision prevention system," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, May 2003, pp. 788–791.

[17] ——, "A 375 × 365 3D 1 k frame/s range-finding image sensor with 394.5 kHz access rate and 0.2 Sub-pixel accuracy," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2004, pp. 118–119.

Yusuke Oike (S'00) was born in Tokyo, Japan, on July 4, 1977. He received the B.S. and M.S. degrees in electronic engineering from the University of Tokyo in 2000 and 2002, respectively. He is currently pursuing the Ph.D. degree in the Department of Electronic Engineering, University of Tokyo.

His current research interests include architecture and design of smart image sensors, mixed-signal circuits, and functional memories.

Mr. Oike is a student member of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICEJ) and The Institute of Image Information and Television Engineers of Japan (ITEJ). He has received the Best Design Awards from IEEE International Conference on VLSI Design 2002 and IEEE ASP-DAC 2004.

Makoto Ikeda (M'99) received the B.S., M.S., and Ph.D. degrees in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991, 1993, and 1996, respectively.

He joined the Electronic Engineering Department, University of Tokyo, as a Faculty Member in 1996, and he is currently an Associate Professor in the VLSI Design and Education Center, University of Tokyo. His interests include the reliability of VLSI design.

Dr. Ikeda is a member of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICEJ) and the Information Processing Society of Japan (IPSJ).

Kunihiro Asada (S'77–M'80) was born in Fukui, Japan, on June 16, 1952. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1975, 1977, and 1980, respectively.

In 1980, he joined the Faculty of Engineering, University of Tokyo, and became a Lecturer, then an Associate Professor, and, finally, a Professor in 1981, 1985, and 1995, respectively. From 1985 to 1986, he was a Visiting Scholar with Edinburgh University, Edinburgh, U.K., supported by the British Council. From 1990 to 1992, he served as the first Editor of the English version of the Institute of Electronics, Information and Communication Engineers of Japan's (IEICE) Transactions on Electronics. In 1996, he established the VLSI Design and Education Center (VDEC) with his colleagues at the University of Tokyo. It is a center supported by the Government to promote education and research of VLSI design in all of the universities and colleges in Japan. He is currently in charge of the head of VDEC. His research interests are in design and evaluation of integrated systems and component devices. He has published more than 400 technical papers in journals and conference proceedings.

Dr. Asada is a member of the IEICE and the Institute of Electrical Engineers of Japan (IEEJ). He has received Best Paper Awards from the IEEJ, the IEICE, and IEEE ICMTS1998.

A CMOS Smart Temperature Sensor With a 3σ Inaccuracy of $\pm 0.5\,^{\circ}\mathrm{C}$ From $-50\,^{\circ}\mathrm{C}$ to $120\,^{\circ}\mathrm{C}$

Michiel A. P. Pertijs, *Student Member, IEEE*, Andrea Niederkorn, Xu Ma, Bill McKillop, *Member, IEEE*, Anton Bakker, *Senior Member, IEEE*, and Johan H. Huijsing, *Fellow, IEEE*

***Abstract*—A low-cost temperature sensor with on-chip sigma-delta ADC and digital bus interface was realized in a 0.5 μm CMOS process. Substrate pnp transistors are used for temperature sensing and for generating the ADC's reference voltage. To obtain a high initial accuracy in the readout circuitry, chopper amplifiers and dynamic element matching are used. High linearity is obtained by using second-order curvature correction. With these measures, the sensor's temperature error is dominated by spread on the base-emitter voltage of the pnp transistors. This is trimmed after packaging by comparing the sensor's output with the die temperature measured using an extra on-chip calibration transistor. Compared to traditional calibration techniques, this procedure is much faster and therefore reduces production costs. The sensor is accurate to within $\pm 0.5\,^{\circ}\mathrm{C}$ (3σ) from $-50\,^{\circ}\mathrm{C}$ to $120\,^{\circ}\mathrm{C}$.**

***Index Terms*—Calibration, curvature correction, dynamic offset cancellation, smart sensors, temperature sensors.**

I. Introduction

INTEGRATED temperature sensors with an on-chip analog-to-digital converter and bus interface find growing application in thermal management systems. These so-called "smart" temperature sensors are widely applied in PCs and laptops to monitor the temperature of the microprocessor, the case, and power-consuming peripheral ICs. This application requires low-cost temperature sensors with a desired inaccuracy below $\pm 1.0\,^{\circ}\mathrm{C}$ [1].

Previous smart temperature sensors were usually calibrated at one fixed temperature, at which their inaccuracy could be trimmed below $\pm 1.0\,^{\circ}\mathrm{C}$ at the cost of a time-consuming (and therefore expensive) calibration after packaging. Their inaccuracy over the industrial temperature range is however larger than $\pm 1.0\,^{\circ}\mathrm{C}$ [2]–[8].

This paper describes in detail a smart temperature sensor which achieves an inaccuracy of $\pm 0.5\,^{\circ}\mathrm{C}$ (3σ) from $-50\,^{\circ}\mathrm{C}$ to $120\,^{\circ}\mathrm{C}$ [9]. Costs are kept low by using a mature 0.5-μm CMOS process and a fast calibration procedure. After packaging, the sensor is calibrated by measuring its die temperature using an extra on-chip calibration transistor. Thus, the required calibration time is greatly reduced compared to a traditional calibration with an external reference thermometer. To obtain a high initial accuracy, dynamic offset cancellation and dynamic element matching are applied in the analog front-end. Good linearity over a wide temperature range is obtained by applying second-order curvature correction.

This paper is organized as follows. Section II introduces the measurement principle, including the curvature correction technique. In Section III, the analog front-end circuitry is discussed, which generates two temperature-dependent currents. These are input to a second-order sigma-delta ADC, which is described in Section IV. The calibration technique is detailed in Section V. The paper ends with experimental results in Section VI and conclusions.

Manuscript received February 5, 2004; revised June 17, 2004. This work was supported by the Dutch Technology Foundation STW.

M. A. P. Pertijs and J. H. Huijsing are with the Electronic Instrumentation Laboratory, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: pertijs@ieee.org).

A. Niederkorn is with Zoran Corporation, Mesa, AZ 85210 USA.

X. Ma is with Microchip Technology, Chandler, AZ 85224 USA.

B. McKillop is with On Semiconductor, Phoenix, AZ 85082 USA.

A. Bakker is with Analog Devices, San Jose, CA 95134 USA.

Digital Object Identifier 10.1109/JSSC.2004.841013

II. Measurement Principle

To convert temperature to a digital value, both a well-defined temperature-dependent signal and a temperature-independent reference signal are required. Both can be derived from the base-emitter voltage of a bipolar transistor, in the form of the thermal voltage kT/q and the silicon bandgap voltage [10]. In a CMOS process, substrate pnp transistors are mostly used for this purpose [11]. These are vertical bipolar transistors with a p^+ diffusion as emitter, an n-well as base, and the p^- substrate as collector.

Two voltages are of interest: the base-emitter voltage V_{BE} of a single transistor in its forward-active region, and the difference ΔV_{BE} between the base-emitter voltages of two such transistors biased at different collector current densities.

A. Temperature Dependence of V_{BE}

From the well-known exponential relation between the collector current I_C and the base-emitter voltage V_{BE}, the following expression for V_{BE} as a function of absolute temperature T can be derived [10]:

$$V_{BE}(T) = V_{g0}\left(1 - \frac{T}{T_r}\right) + \frac{T}{T_r}V_{BE}(T_r) - \eta\frac{kT}{q}\ln\left(\frac{T}{T_r}\right) + \frac{kT}{q}\ln\left(\frac{I_C(T)}{I_C(T_r)}\right) \quad (1)$$

where V_{g0} is the extrapolated bandgap voltage at 0 K, η is a process-dependent constant, k is Boltzmann's constant, q is the electron charge, and T_r is an arbitrary reference temperature. As illustrated in Fig. 1(a), $V_{BE}(T)$ is an almost linear function of temperature, with a typical slope of -2 mV/K. The nonlinearity, or curvature, is represented by the last two terms of (1).

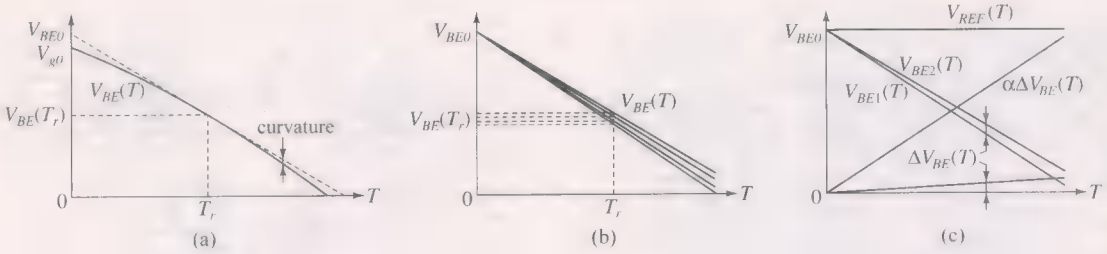


Fig. 1. (a) Temperature dependence of the base-emitter voltage V_{BE}. (b) Variation in V_{BE} due to process spread (curvature omitted for clarity). (c) Combination of V_{BE} and ΔV_{BE} to yield the bandgap reference voltage V_{REF} (curvature again omitted).

It depends on the constant η and on the temperature dependence of the collector current.

The slope of the base-emitter voltage depends on process parameters and the absolute value of the collector current. Its extrapolated value at 0 K, however, is insensitive to process spread and current level, as illustrated in Fig. 1(b). Therefore, a calibration at one temperature can be used to trim the slope of V_{BE} to a desired value [12].

V_{BE} is also sensitive to stress. Fortunately, substrate pnp transistors are much less stress-sensitive than other bipolar transistors [13]. Packaging-induced shifts in V_{BE} will be corrected by calibrating the sensor after packaging, as will be discussed in Section V.

B. Temperature Dependence of ΔV_{BE}

The difference ΔV_{BE} between the base-emitter voltages of a transistor operated at two collectors I_{C1} and I_{C2} can be expressed as [10]

$$\Delta V_{BE}(T) = V_{BE2}(T) - V_{BE1}(T) = \frac{kT}{q} \ln\left(\frac{I_{C2}}{I_{C1}}\right). \quad (2)$$

Provided the collector-current ratio is constant, ΔV_{BE} is proportional to absolute temperature (PTAT), as shown in Fig. 1(c).

In contrast with V_{BE}, ΔV_{BE} is independent of process parameters and the absolute value of the collector currents.[1] Moreover, it is insensitive to stress [15]. Its temperature coefficient is, however, typically an order of magnitude smaller than that of V_{BE} (depending on the collector current ratio).

C. Combining V_{BE} and ΔV_{BE}

In a bandgap voltage reference, an amplified version of ΔV_{BE} is added to V_{BE} to yield a temperature-independent reference voltage V_{REF}, as illustrated in Fig. 1(c). In our temperature sensor, this addition is implemented in the current domain at the input of the sigma-delta modulator (Fig. 2). Depending on the bitstream output bs of the modulator, either a current $\Delta V_{BE}/R_1$ is integrated (when $bs = 0$) or a current $-V_{BEtrim}/R_2$ (when $bs = 1$), where V_{BEtrim} is a trimmed base-emitter voltage. The negative feedback in the modulator

[1]Often a multiplicative factor n is included in the equation for ΔV_{BE} to model the influence of the reverse Early effect and other nonidealities [14]. If V_{BE} and ΔV_{BE} are generated using transistors biased at approximately the same current density, an equal multiplicative factor will appear in V_{BE}. In a smart temperature sensor, these factors cancel, and will therefore not be considered further.

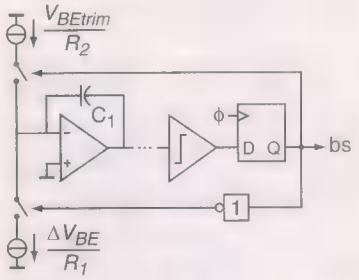


Fig. 2. Simplified circuit diagram of the sigma-delta modulator.

will ensure that the average current flowing into the integrator is zero. This implies

$$\mu \frac{V_{BEtrim}}{R_2} = (1-\mu)\frac{\Delta V_{BE}}{R_1}$$
$$\Longrightarrow \mu = \frac{\alpha \Delta V_{BE}}{V_{BEtrim} + \alpha \Delta V_{BE}} = \frac{\alpha \Delta V_{BE}}{V_{REF}} \quad (3)$$

where μ is the average value of the bitstream (i.e., the fraction of 1's), and $\alpha = R_2/R_1$. The denominator of (3) is essentially a bandgap reference voltage, while the numerator is PTAT. The average μ will therefore also be PTAT, so that the bitstream can be used, with appropriate scaling in the digital decimation filter, to produce a digital representation of the chip's temperature in degrees Celsius.

With the configuration of Fig. 2, only about 30% of the dynamic range of the sigma-delta modulator is used, since $\mu = 0$ corresponds to -273 °C and $\mu = 1$ corresponds to approximately 325 °C, while the temperature range of interest is from -50 °C to 125 °C. Other combinations of V_{BEtrim} and ΔV_{BE} can be used to utilize more of the dynamic range [16], but these require copying or scaling of the currents, thus introducing more sources of errors. Since a second-order sigma-delta modulator is used, which can easily provide sufficient resolution, a more efficient use of the dynamic range is not needed. In fact, for the single-loop modulator used (Section IV), the quantization noise strongly increases for μ close to 0 or 1. With the configuration of Fig. 2, these regions are conveniently avoided.

D. Curvature Correction

The curvature of V_{BE} will also be present in the reference voltage V_{REF}, which, in turn, results in a nonlinearity in $\mu(T)$. The curvature is modeled by the last two terms in (1). For a value of $\eta = 4.4$ for our process and a PTAT collector current (as used

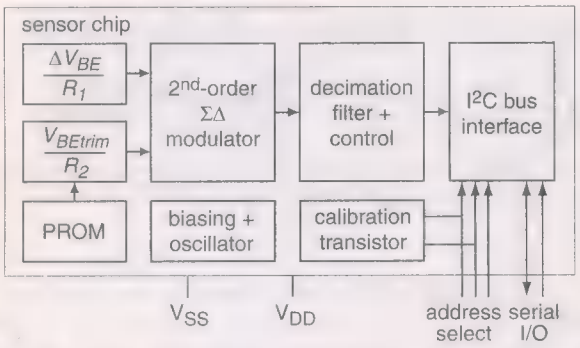


Fig. 3. Block diagram of the temperature sensor.

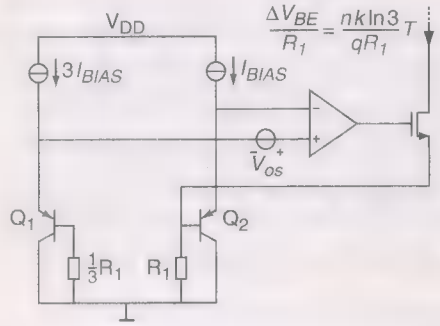


Fig. 4. Simplified circuit diagram of the ΔV_{BE}-dependent current source.

in our design), the corresponding nonlinearity amounts to 2 °C over the temperature range of −50 °C to 125 °C.

Fortunately, the second-order component of the curvature can easily be eliminated by giving V_{REF} a small positive temperature coefficient [4], [17], i.e., by making α in (3) slightly larger than in a bandgap reference. With an appropriate value for α (22 in our case), such a temperature-dependent V_{REF} gives rise to a second-order nonlinearity in $\mu(T)$ which exactly cancels the second-order nonlinearity originating from V_{BE}. What remains is a third-order nonlinearity of about 0.3 °C over the temperature range.

E. Block Diagram

The input currents for the sigma-delta modulator of Fig. 2 are generated by a $\Delta V_{BE}/R_1$ current source and a V_{BEtrim}/R_2 current source, as shown in the block diagram of Fig. 3. A decimation filter converts the bitstream output of the modulator to a digital representation of the temperature, also taking care of the scaling required to convert the average value μ of the bitstream to °C. The result is communicated to the outside world using an I^2C bus interface. Also on the chip are the calibration transistor, a PROM to hold the setting of the trimming of V_{BE}, a biasing circuit and an oscillator.

III. Temperature-Dependent Current Sources

A. ΔV_{BE}-Dependent Current Source

A simplified circuit diagram of the $\Delta V_{BE}/R_1$ current source is shown in Fig. 4 [16]. Two substrate pnp transistors Q_1 and Q_2 are biased at a 3:1 current ratio. The bias currents are generated in a separate circuit (not shown). The resulting difference in base-emitter voltage ΔV_{BE} has a sensitivity of 100 μV/°C. By

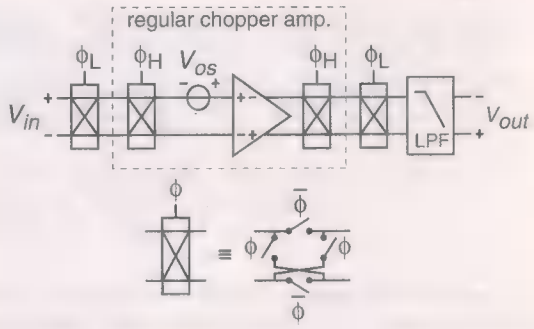


Fig. 5. Principle of a nested-chopper amplifier.

means of the feedback loop, ΔV_{BE} is generated across a resistor R_1 in series with the base of Q_2, resulting in the desired output current. To avoid that the output current is affected by the base current of Q_2, a resistor $R_1/3$ is added in series with the base of Q_1. As the base current of Q_1 is three times as large as that of Q_2, the base currents result in an equal voltage drop across both resistors, which is a small common-mode change that does not affect the output current.

The inaccuracy of the circuit of Fig. 4 is mainly determined by the offset V_{os} of the opamp, which directly adds to ΔV_{BE}. To result in a negligible temperature error (0.1 °C), this offset has to be smaller than 10 μV. Since typical offsets of CMOS opamps are in the millivolt range, offset cancellation is required. Mismatch in the current sources or the pnp transistors also leads to temperature errors. For these errors to be negligible, the matching has to be better than 0.035%, which requires dynamic element matching.

The offset of the opamp can be reduced using the chopping technique. In a regular chopper amplifier, a pair of chopper switches is added around the amplifier whose offset V_{os} needs to be cancelled (Fig. 5) [16]. The chopper at the input modulates the input signal to the frequency of control signal ϕ_H, which lies above the offset and $1/f$ corner frequency of the amplifier. The chopper at the output demodulates the amplified input signal, and simultaneously modulates the amplified offset and $1/f$ noise to the frequency of ϕ_H, where they can be filtered out by a low-pass filter (LPF).

Due to charge injection and clock feedthrough, a regular chopper amplifier has a typical residual offset of a few tens of microvolts. To reduce the offset below 10 μV, an extra outer pair of chopper switches is added. This is controlled by a low-frequency control signal ϕ_L. This pair modulates the regular chopper amplifier's residual offset to the frequency of ϕ_L, where it can also be removed by the LPF. The residual offset of the resulting *nested*-chopper amplifier is determined by clock feedthrough and charge injection in the low-frequency chopper switches, and is therefore much smaller than that of the regular chopper amplifier. Residual offsets as low as 100 nV have been reported [18].

Fig. 6 shows how the nested-chopper amplifier is embedded in the $\Delta V_{BE}/R_1$ current source. The opamp is split up into three stages, with chopper switches between them. The first stage is a folded-cascode amplifier, the second stage is a differential pair, and the third stage is its current mirror load. Miller compensation (not shown) is used to stabilize the opamp.

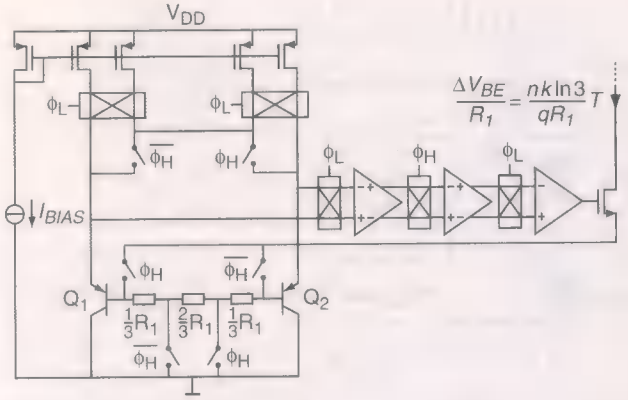


Fig. 6. Detailed circuit diagram of the ΔV_{BE}-dependent current source.

The input chopper driven by ϕ_H is implemented in the current domain, by switching between a 3:1 and 1:3 current ratio. Thus, offset resulting from mismatch between the pnp transistors is also chopped. To maintain the correct feedback polarity, the connection to the output transistor is switched back and forth between the bases of Q_1 and Q_2. As in Fig. 4, compensation for the base currents is realized by making sure that a resistor $R_1/3$ is in series with the base of the transistor that carries the larger bias current.

The bias currents are generated by four current sources of 0.5 μA each, which are dynamically matched using the control signals ϕ_H and ϕ_L. Alternately, one of the current sources biases one transistor, while the remaining three bias the other.

The control signal ϕ_H switches at 16 kHz, while ϕ_L switches at 80 Hz. The modulated offset and $1/f$ noise components are filtered out by the sigma-delta modulator and the decimation filter, as will be discussed in Section IV.

B. V_{BE}-Dependent Current Source

The trimmed base-emitter voltage V_{BEtrim} is generated by adjusting the base-emitter voltage V_{BE} of a substrate pnp transistor with a small programmable PTAT voltage. Fig. 7 shows how this is implemented: a PTAT current is passed through a digitally programmable resistor in series with a diode-connected substrate pnp. The PTAT voltage across this resistor compensates for the PTAT-type spread on V_{BE} [Fig. 1(b)]. The PTAT current in Fig. 7 is generated in a separate bias circuit (not shown).

The current V_{BEtrim}/R_2 is generated using a voltage-to-current converter around a regular chopper amplifier controlled by ϕ_L. Because of the higher sensitivity of $V_{BE}(-2\,\text{mV}/^{\circ}\text{C})$, a nested-chopper amplifier was not needed here. The amplifier has a folded-cascode topology. To accurately define the ratio α in (3), the resistors R_1 and R_2 are made of identical unit resistors.

To save power, the nominal output current is kept relatively small (0.5 μA). Therefore, a large resistance (more than 1 MΩ) is required. In order to reduce the size of the resistor, a current mirror with a dynamically matched 3:1 ratio is used. The dynamic element matching is again controlled by ϕ_L and ϕ_H. Thus, the chip area required for the resistor is reduced by a factor 3 without using special high-resistivity resistors (which would require extra processing steps).

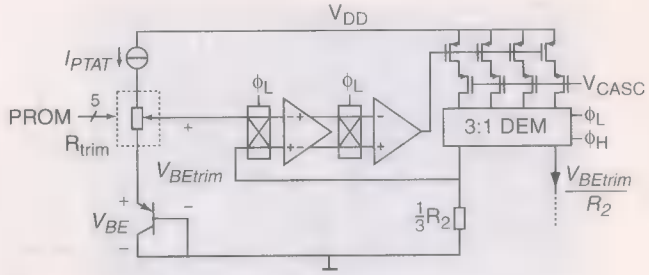


Fig. 7. Circuit diagram of the V_{BE}-dependent current source.

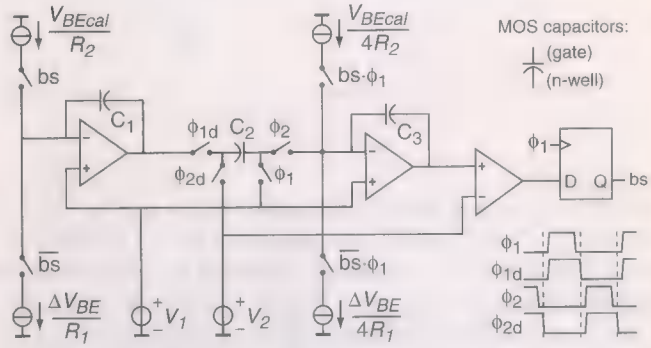


Fig. 8. Circuit diagram of the sigma-delta modulator; initialization circuits are omitted for clarity; unused currents are switched to V_1.

IV. Sigma-Delta ADC

A sigma-delta ADC is used to convert the temperature-dependent currents into a digital temperature reading. A quantization noise below 0.05 °C in a conversion time of 30 ms was desired. With a first-order sigma-delta modulator, as was used in previous work [4], [16], this would require a clock frequency of about 500 kHz. As this would lead to an undesirably high power consumption, a second-order modulator was used, which requires a clock frequency of only 16 kHz.

As in an incremental ADC [19], the integrators of the modulator are reset at the beginning of the conversion, and a second-order decimation filter is used rather than the usual third-order filter. In contrast with an incremental ADC, however, the input signal is not sampled and held during the conversion, but it is integrated continuously so as to filter out the modulated offset and $1/f$ noise.

A. Sigma-Delta Modulator

A simplified circuit diagram of the sigma-delta modulator is shown in Fig. 8. It is clocked using a nonoverlapping clock which runs at the same frequency as the control signal ϕ_H in the current sources. This ensures that modulated offset at harmonics of ϕ_H is averaged out within a clock cycle of the modulator. As discussed in Section II-C, the bitstream determines which of the two input currents is integrated on the first integrator. Unused currents are dumped into a reference node at V_1 (not shown).

During clock phase ϕ_1, the output of the first integrator is sampled on capacitor C_2. During phase ϕ_2, the charge is transferred to the second integrator, the output of which is fed into a clocked comparator that produces the bitstream bs. To minimize charge injection onto C_2, clock signals ϕ_{1d} and ϕ_{2d} have

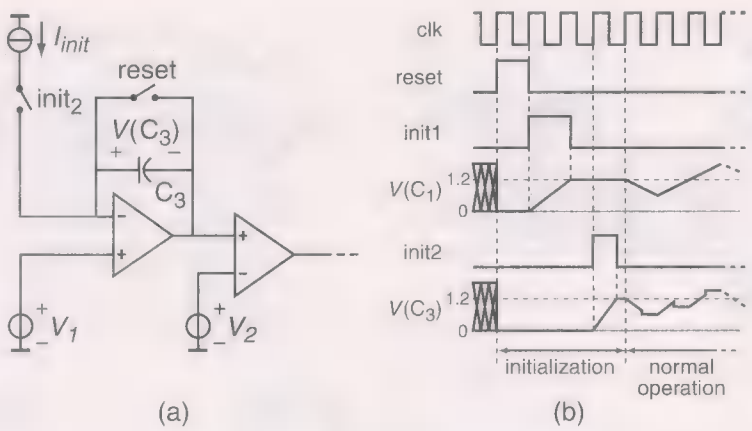


Fig. 9. (a) Initialization circuit for the second integrator. (b) Waveforms during the initialization sequence.

delayed downgoing edges with respect to ϕ_1 and ϕ_2. Scaled copies of the input currents are integrated on the second integrator during phase ϕ_1 to ensure stability of the modulator and to minimize the swing at the output of the first integrator. The scaled copies are not critical for the dc accuracy of the modulator; mismatches up to several percent can be tolerated.

The modulator is implemented using MOS capacitors, to avoid the extra processing steps required for linear capacitors. C_1 and C_2 are made from identical unit capacitors to ensure linear charge transfer in spite of the nonlinearity of these capacitors. The nonlinearity of C_3 is not relevant, since only the sign of the output of the second integrator is detected by the comparator.

To maximize the capacitance per area of the MOS capacitors, and to avoid operating them in their most nonlinear region (around 0 V), they are biased in accumulation. The gates of the capacitors are at V_1, while the feedback ensures that the average voltage on their wells is V_2. Therefore, they can be biased in accumulation by choosing V_1 sufficiently higher than V_2 (1.2 V in this case).

B. Initialization Sequence

In contrast with the usual continuous operation of sigma-delta ADCs, the temperature sensor requires a "one-shot" type of operation, i.e., the converter is powered up, produces a single conversion result, and powers down again to save power. This has implications for both the initialization of the modulator, and the design of the decimation filter.

After power-up, the modulator is brought into a well-defined state by resetting the integration capacitors. After the reset, the integration capacitors could be driven into accumulation by the feedback loop, but this may take many clock cycles (depending on the input signal). To expedite this, the capacitors are precharged using an initialization current I_{init}, as shown in Fig. 9(a) for the second integrator. The initialization current is switched to the input of the integrator until its output reaches the voltage V_2, which is detected by the comparator.

To allow for similar initialization of the first integrator, its output can be connected to the input of the comparator using a set of switches (not shown). The total initialization sequence consists of resetting both integration capacitors, precharging the capacitor of first integrator, and then precharging that of the second integrator. The corresponding waveforms are shown in Fig. 9(b).

C. Decimation Filter

Once the modulator has reached its steady state, the bitstream is fed into a decimation filter, which produces a single conversion result. Usually, the order of a sinc decimation filter is chosen one higher than that of the loop filter [20], which implies a third-order filter for our second-order modulator. However, for a given conversion time, and thus a given impulse response length of the filter, the corner frequency of a third-order filter is higher than that of a second-order filter. Due to this higher corner frequency, the use of a third-order filter will result in more quantization noise, in spite of its faster roll-off. Therefore, a less complex sinc^2 filter is used rather than a sinc^3 filter.

For the chopping and dynamic element matching in the current sources to be effective, the decimation filter has to filter out the residuals modulated by the low-frequency control signal ϕ_L. Therefore, ϕ_L is clocked at a frequency that coincides with the first zero in the frequency response of the sinc^2 filter, which is at approximately 80 Hz.

The decimation filter is implemented by an up/down counter and an accumulator. The counter counts up during the first half of the decimation period and down during the second half, thus realizing the triangular impulse response of a sinc^2 filter. The accumulator adds the counter value if the bitstream is "1". The initial value of the accumulator and the exact length of the decimation period (and thereby the gain of the filter) are chosen such that the accumulated value at the end of the conversion can be directly interpreted as a temperature in degrees Celsius.

V. Calibration Technique

To calibrate any integrated temperature sensor, its temperature reading has to be compared to that of a reference thermometer at the same temperature as the sensor chip. The difference between the readings may then be used to trim the sensor. This calibration is often done at wafer-level, which has the advantage that the temperature of the whole wafer can be stabilized

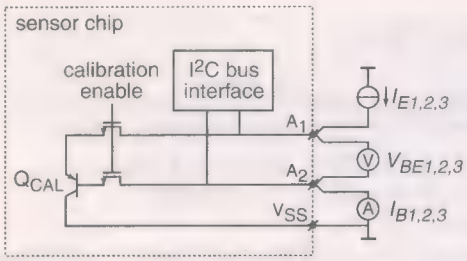


Fig. 10. Connection of the calibration transistor by reusing digital input pins.

Fig. 11. Chip micrograph of the temperature sensor.

and measured, after which the individual sensors can be calibrated and trimmed using a wafer prober. An important disadvantage of this approach, however, is that additional errors introduced by packaging stress are not taken into account. Even when the sensor design is based on relatively stress-insensitive substrate pnp transistors, a significant error will result if a low-cost plastic package is used. Experiments on a bandgap reference based on such pnps have shown shifts up to 2 mV in V_{BE} [13]. As can be derived from (3), this translates to a temperature error of about 0.5 °C. Therefore, it is desired to do the calibration after packaging.

If the temperature of each individual packaged sensor has to be stabilized and measured with an inaccuracy below $\pm 0.5\,^{\circ}$C using a reference thermometer, this becomes the dominant contributor to the test time of the sensor. A faster and therefore cheaper alternative is to make use of the process- and stress-insensitivity of ΔV_{BE} (discussed in Section II-B): an extra substrate pnp transistor, the calibration transistor, has been integrated on the sensor chip, and is used as a reference thermometer *inside* the package [21]. From its ΔV_{BE}, measured using external electronics, the die temperature can be determined within $\pm 0.1\,^{\circ}$C [21]. As it is integrated on the same thermally conducting silicon as the sensor circuit, very little thermal settling time is required. Moreover, the requirements on the thermal stability of the production setup are relaxed.

Fig. 10 shows how the calibration transistor (Q_{CAL}) is connected without reserving extra pins for it. Two existing address pins of the I²C bus interface are reused during calibration to connect to the base and emitter of Q_{CAL}. During normal operation, Q_{CAL} is isolated from these pins using MOS switches. These switches are controlled via the bus interface.

The temperature of the on-chip calibration transistor is determined by applying a number of bias currents to it, and measuring its base-emitter voltage and base current using external electronics. Thus, ΔV_{BE} can be measured while compensating for series resistances [22] and current-gain variations. From this, the chip temperature can be calculated with an absolute accuracy of $\pm 0.1\,^{\circ}$C [21]. The difference between this temperature and a reading of the sensor is then used to determine the appropriate setting for the programmable resistor R_{trim} in Fig. 7. This setting is then programmed in PROM via the I²C bus interface.

VI. Experimental Results

The temperature sensor was fabricated in a standard 0.5-μm CMOS process. A chip micrograph is shown in Fig. 11. The chip area is 2.5 mm², of which about half is used for the digital bus interface and control.

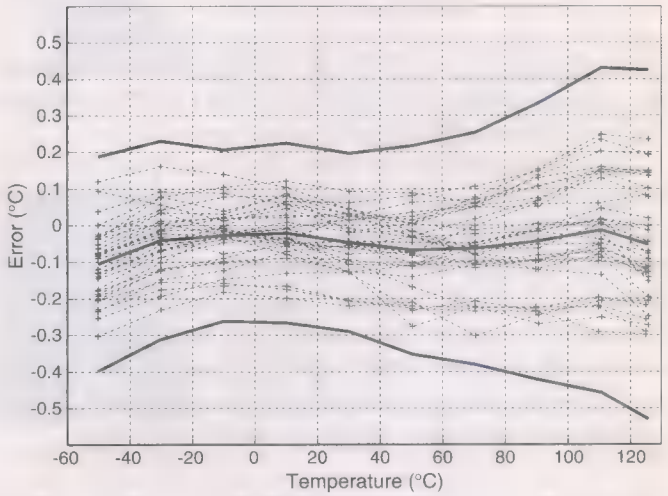


Fig. 12. Measured temperature error of 32 samples from one batch, with average and $\pm 3\sigma$ values.

TABLE I
Performance Summary

Technology	0.5μm CMOS
Chip size	2.5mm²
Supply voltage	2.7V – 5.5V
Temperature range	-50°C – 125°C
Conversion rate	0.125 – 30 conversions/s
Noise level	0.03°C$_{rms}$
Supply current	130μA at 10 conversions/s
Power supply rejection	0.3°C/V from 3.0V to 3.6V
Inaccuracy (3σ)	$\pm$0.3°C at 25°C $\pm$0.5°C from -50°C to 120°C

Fig. 12 shows the measured temperature error of 32 samples from one processing batch, operated at a supply voltage of 3.3 V. These samples were packaged in 8-pin ceramic packages. They were calibrated and trimmed at room temperature using the described procedure, after which they were placed in an oven along

TABLE II
COMPARISON OF INACCURACY WITH PREVIOUS WORK

Reference	Inaccuracy	Range	Conditions	Calibration
Bakker, 1996 [2]	±1.0°C	−40°C to 120°C	min/max of 3 samples	after packaging, 2 points
Tuthill, 1998 [3]	±1.5°C	−50°C to 125°C	min/max of 6 samples	wafer-level, 1 point
Pertijs, 2001 [4]	±1.5°C	−50°C to 125°C	±3σ of 32 samples	batch-calibration
LM92 [5]	±0.33°C	30°C	min/max	unknown
	±1.5°C	−25°C to 150°C	min/max	unknown
DS1626 [6], ADT7301 [7]	±0.5°C	0°C to 70°C	min/max	unknown
	±2.0°C	−55°C to 125°C	min/max	unknown
SMT160-30 [8]	±0.7°C	−30°C to 100°C	min/max	wafer-level, 1 point
	±1.2°C	−45°C to 130°C	min/max	wafer-level, 1 point
This work	±0.3°C	25°C	±3σ of 32 samples	after packaging, 1 point
	±0.5°C	−50°C to 125°C	±3σ of 32 samples	after packaging, 1 point

with a platinum resistor calibrated to 20 mK. Their 3σ inaccuracy in the temperature range of −50 °C to 120 °C is ±0.5 °C. The performance of the chips is summarized in Table I.

Table II compares the inaccuracy with that of previous work. Though many smart temperature sensors have been published, only a few publications provide sufficient measurement results for a proper comparison [2]–[4]. Since most work in this field is done in industry, the inaccuracy specifications of four leading commercial temperature sensors have also been included in the table [5]–[8]. At room temperature, the presented sensor performs as well as the best-performing previous work, while over a wide temperature range it performs significantly better.

VII. CONCLUSION

A CMOS temperature sensor with integrated second-order sigma-delta ADC and bus interface has been presented. A high initial accuracy is achieved by applying dynamic offset cancellation and dynamic element matching in the front-end circuitry, and by applying a linearization technique that eliminates the second-order curvature. With these measures, the spread on the base-emitter voltage is the dominant source of errors. This is trimmed based on the results of a single-point calibration, which takes place after packaging. The chip temperature is determined from the electrical characteristics of an additional on-chip transistor, which are measured using external electronics. Thus a fast and accurate calibration can be performed. After calibration at room temperature and trimming, the sensor has a 3σ inaccuracy of ±0.5 °C in the temperature range of −50 °C to 120 °C, which is, to date, the highest reported accuracy for this type of sensors.

ACKNOWLEDGMENT

The authors would like to thank Philips Semiconductors for the fabrication of the sensors.

REFERENCES

[1] D. Marsh, "Silicon sensors harness thermal management," *EDN*, pp. 43–55, Dec. 2003.

[2] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE J. Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, Jul. 1996.

[3] M. Tuthill, "A switched-current, switched-capacitor temperature sensor in 0.6-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 33, no. 7, pp. 1117–1122, Jul. 1998.

[4] M. A. P. Pertijs, A. Bakker, and J. H. Huijsing, "A high-accuracy temperature sensor with second-order curvature correction and digital bus interface," in *Proc. ISCAS*, May 2001, pp. 368–371.

[5] (2000) LM92 Data Sheet. National Semiconductor Corporation. [Online]. Available: http://www.national.com

[6] (2003) DS1626 Data Sheet. Maxim Int. Products. [Online]. Available: http://www.maxim-ic.com

[7] (2003) ADT7301 Data Sheet. Analog Devices Inc. [Online]. Available: http://www.analog.com

[8] (2003) SMT160-30 Data Sheet. Smartec B.V. [Online]. Available: http://www.smartec.nl

[9] M. Pertijs *et al.*, "A CMOS temperature sensor with a 3σ inaccuracy of ±0.5 °C from −50 °C to 120 °C," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2003, pp. 200–201, 488.

[10] G. C. M. Meijer, "Thermal sensors based on transistors," *Sensors and Actuators*, vol. 10, pp. 103–125, Sep. 1986.

[11] G. Wang and G. C. M. Meijer, "Temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators A*, vol. 87, pp. 81–89, Dec. 2000.

[12] G. C. M. Meijer, G. Wang, and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," *IEEE Sensors J.*, vol. 1, no. 3, pp. 225–234, Oct. 2001.

[13] F. Fruett, G. C. M. Meijer, and A. Bakker, "Minimization of the mechanical-stress-induced inaccuracy in bandgap voltage references," *IEEE J. Solid-State Circuits*, vol. 38, no. 7, pp. 1288–1291, Jul. 2003.

[14] M. A. P. Pertijs, G. C. M. Meijer, and J. H. Huijsing, "Precision temperature measurement using CMOS substrate pnp transistors," *IEEE Sensors J.*, vol. 4, no. 3, pp. 294–300, Jun. 2004.

[15] F. Fruett, G. Wang, and G. C. M. Meijer, "The piezojunction effect in npn and pnp vertical transistors and its influence on silicon temperature sensors," *Sensors and Actuators A*, vol. 85, pp. 70–74, Aug. 2000.

[16] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*, ser. Int. Series in Engineering and Computer Science. Boston, MA: Kluwer, 2000, vol. 595.

[17] G. C. M. Meijer *et al.*, "A three-terminal integrated temperature transducer with microcomputer interfacing," *Sensors and Actuators*, vol. 18, pp. 195–206, Jun. 1989.

[18] A. Bakker, K. Thiele, and J. H. Huijsing, "A CMOS nested-chopper instrumentation amplifier with 100-nV offset," *IEEE J. Solid-State Circuits*, vol. 35, no. 12, pp. 1877–1883, Dec. 2000.

[19] J. Robert and P. Deval, "A second-order high-resolution incremental A/D converter with offset and charge injection compensation," *IEEE J. Solid-State Circuits*, vol. SSC-23, no. 3, pp. 736–741, Jun. 1988.

[20] S. R. Norsworthy, R. Schreier, and G. C. Temes, *Delta-Sigma Data Converters: Theory, Design and Simulation*. Piscataway, NJ: IEEE Press, 1997.

[21] M. A. P. Pertijs and J. H. Huijsing, "Transistor temperature measurement for calibration of integrated temperature sensors," in *Proc. IMTC*, May 2002, pp. 755–758.

[22] J. M. Audy and B. Gilbert, "Multiple sequential excitation temperature sensing method and apparatus," U.S. Patent 5,195,827, Mar. 4, 1993.

Michiel A. P. Pertijs (S'99) was born in Roosendaal, The Netherlands, on May 31, 1977. He received the M.Sc. degree in electrical engineering (*cum laude*) from Delft University of Technology in 2000. He is currently working toward the Ph.D. degree at the Electronic Instrumentation Laboratory of the same university, on the subject of high-accuracy CMOS smart temperature sensors.

In 2000, he was an intern with Philips Semiconductors, Sunnyvale, CA, working on analog circuit design. From 1997 to 1999, he worked part-time for EARS B.V., Delft, The Netherlands, on the production and development of a handheld photosynthesis meter. His research interests include analog and mixed-signal interface electronics and smart sensors.

Andrea Niederkorn was born on November 14, 1969, in San Diego, CA. She received the B.S.E.E. degree from the University of Texas, El Paso, in 1991, and the M.S.E.E. degree from Arizona State University, Tempe, in 1996.

Her professional experience has been with electronic sensors, and she is currently with Zoran Corporation developing CMOS image sensors for the mobile communications market.

Xu Ma was born in Beijing, China. He received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, in 1996 and 1999, respectively. He received a second M.S. degree in solid-state circuit design from Arizona State University, Tempe, in 2002.

He joined Philips Semiconductors North America, Tempe, in 2000, where he was engaged in the development of high-accuracy temperature sensor project. He joined Microchip Technology Inc., Chandler, AZ, in 2003. His current interests include micro-controller and mixed- signal circuit design.

Bill McKillop (M'01) received the B.S.E.E. and M.S.E.E. degrees from the University of Arizona, Tucson, in 1997 and 2000.

In 1999, he worked as a Design Engineer in the Magnetic Storage Division for Cirrus Logic, Austin, TX. From 2000 to 2004, he worked as a Design Engineer for the Standard Analog Group for Philips, Tempe, AZ, specializing in power management. He is currently with On Semiconductor, Phoenix, AZ, in the Amplifier Group.

Anton Bakker (S'95–M'00–SM'01) received the M.Sc. and Ph.D. degrees in electrical engineering from the Delft University of Technology, The Netherlands, in 1991 and 2000, respectively.

He was an Assistant Professor at Delft University from 1991 to 2000. In 1997, he started as a consultant for Philips, Sunnyvale, CA, and became a Senior Design Engineer for Philips in 2000. In early 2004, he joined Analog Devices as a Senior Staff Design Engineer. He is author or co-author of over 25 international papers and holds five patents. He wrote the book *High-Accuracy CMOS Smart Temperature Sensors* (Kluwer, 2000).

Johan H. Huijsing (SM'81–F'97) was born on May 21, 1938. He received the M.Sc. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 1969, and the Ph.D. degree from the same University in 1981 for his thesis on operational amplifiers.

He has been an Assistant and Associate Professor in electronic instrumentation with the Faculty of Electrical Engineering of the Delft University of Technology since 1969, where he became a full Professor in the Chair of Electronic Instrumentation in 1990, and Professor Emeritus in 2003. From 1982 to 1983, he was a Senior Scientist at Philips Research Laboratories, Sunnyvale, CA. Since 1983, he has been a consultant for Philips Semiconductors, Sunnyvale, and since 1998 also a consultant for Maxim, Sunnyvale. His research work is focused on the systematic analysis and design of operational amplifiers, analog-to-digital converters, and integrated smart sensors. He is author or co-author of some 200 scientific papers, 40 patents, and 9 books, and co-editor of 11 books.

Dr. Huijsing is a Fellow of the IEEE for contributions to the design and analysis of analog integrated circuits. He was awarded the title of Simon Stevin Meester for Applied Research by the Dutch Technology Foundation. He is initiator and co-chairman of the International Workshop on Advances in Analog Circuit Design, which has been held annually since 1992 in Europe. He was a member of the program committee of the European Solid-State Circuits Conference from 1992 to 2002. He was Chairman of the Dutch STW Platform on Sensor Technology and Chairman of the Biennial National Workshop on Sensor Technology from 1991 to 2002.

A Four-Channel 3.125-Gb/s/ch CMOS Serial-Link Transceiver With a Mixed-Mode Adaptive Equalizer

Jinwook Kim, Jeongsik Yang, Sangjin Byun, Hyunduk Jun, Jeongkyu Park, Cormac S. G. Conroy, *Member, IEEE*, and Beomsup Kim, *Fellow, IEEE*

***Abstract*—This paper presents a quad-channel serial-link transceiver providing a maximum full duplex raw data rate of 12.5 Gb/s for a single 10-Gbit eXtended Attachment Unit Interface (XAUI) in a standard 0.18-μm CMOS technology. To achieve low bit-error rate (BER) and high-speed operation, a mixed-mode least-mean-square (LMS) adaptive equalizer and a low-jitter delay-immune clock data recovery (CDR) circuit are used. The transceiver achieves BER lower than $< 4.5 \times 10^{-15}$ while its transmitted data and recovered clock have a low jitter of 46 and 64 ps in peak-to-peak, respectively. The chip consumes 178 mW per each channel at 3.125-Gb/s/ch full duplex (TX/RX simultaneous) data rate from 1.8-V power supply.**

***Index Terms*—Adaptive equalizer, clock data recovery (CDR), serial-link transceiver.**

I. Introduction

IN MODERN electrical interconnect systems, high-speed serial links have replaced parallel data buses, and serial link speed is rapidly increasing due to the evolution of CMOS technology. For example, high-end routers and backbone switches have wide parallel buses to communicate to network terminals such as network processors. High pin counts result in high-cost processors and switches, and makes system engineering and board design difficult because of coupling and skew between bus lines. High-speed serial links eliminate these problems.

Serial link performance is limited by: 1) noise, which introduces timing and amplitude errors, and 2) the bandwidth limitations of the electronic components. In order to resolve the inter-symbol interference (ISI) problems caused by bandwidth limitations, pre-emphasis techniques are used on the transmitter side [2], and adaptive equalization is used on receiver side [3]. Pre-emphasis is good for well-known channel characteristics but it cannot adapt to channel variations. Moreover, the larger voltage swing caused by pre-emphasis generates more ringing.

An adaptive equalizer compensates channel distortion caused by limited bandwidth. Analog implementations have the advantage of filtering speed over digital implementations. Furthermore, even though analog approaches suffer from the nonidealities of analog components and noise, analog approaches have the advantage that as the filtering occurs before sampling, they avoid the signal processing delays—i.e., latency—due to digital filtering, which affect the performance and stability of the phase-locked loop (PLL) that provides the sampling clock [8]. Two kinds of analog implementation have been used. One is a sampling-type equalizer [5] and the other is a continuous-time equalizer [7]. With the sampling-type equalizer, the sample-and-hold circuits become unstable as the data-rate increases. Recent research has introduced a post-equalizer [3] at several Gb/s rates, but without any adaptation algorithm. This paper proposes a mixed-mode adaptive equalizer that takes advantages of both high-speed analog continuous-time filtering and the stability of digital tap adaptation.

One key building block of an analog continuous-time transversal equalizer is an analog delay line. In order to meet the required one bit clock period delay, programmability and tuning circuits are normally necessary. This paper introduces an analog delay line that generates exact 1-bit delay without any tuning circuits.

The clock data recovery (CDR) circuit plays a critical role in the receiver. It extracts the clock and regenerates data from the input data stream and reduces the timing error, one of the critical system performance limiting factors. In low-frequency applications a digital PLL can be used for good jitter suppression or jitter tolerance [9]. Phase-tracking CDRs have been used for several Gb/s rates [14], [15] because they do not suffer from phase quantization errors. Comparing the two kinds of phase detection methods, the binary CDR is more suitable for high-speed operation than the linear CDR because it does not suffer from the timing offset caused by setup/hold-timing uncertainty of the sampler [16].

The jitter of a binary CDR circuit is set by the minimum resolution of the phase interpolator because of its bang-bang operation [6]. In the case of an *ideal* CDR circuit with *no* delay, which immediately updates the timing, the recovered clock jitter is limited by the minimum resolution of the phase interpolator. If there are some delays in the recovery loop, the jitter is more than the minimum resolution because the delays in the recovery loop prevent immediate timing update. In this paper, we present a new delay-immune CDR circuit. By ignoring the successive Up/Dn value of the delay amount in the recovery loop, it can implement an ideal bang-bang operation and reduce the jitter of the recovered clock.

This paper is organized as follows. The structure of the proposed transceiver architecture is presented in Section II.

Manuscript received October 2, 2003; revised June 6, 2004.

J. Kim, H. Jun, and J. Park are with the Berkäna Wireless, Inc., Campbell, CA 95008 USA, and also with Berkäna Korea, Seoul 138-803, Korea (e-mail: jinux@berkanawireless.com; hdjun@berkanawireless.com; asicpark@berkanawireless.com).

J. Yang, C. S. G. Conroy, and B. Kim are with Berkäna Wireless, Inc., Campbell, CA 95008 USA (e-mail: jsyang@berkanawireless.com; cconroy@berkanawireless.com; bkim@berkanawireless.com).

S. Byun is with Berkäna Wireless, Inc., Campbell, CA 95008, USA, and also with the Electronics and Telecommunications Research Institute (ETRI), Daejon 305-350, Korea, (e-mail: sjbyun@etri.re.kr).

Digital Object Identifier 10.1109/JSSC.2004.841037

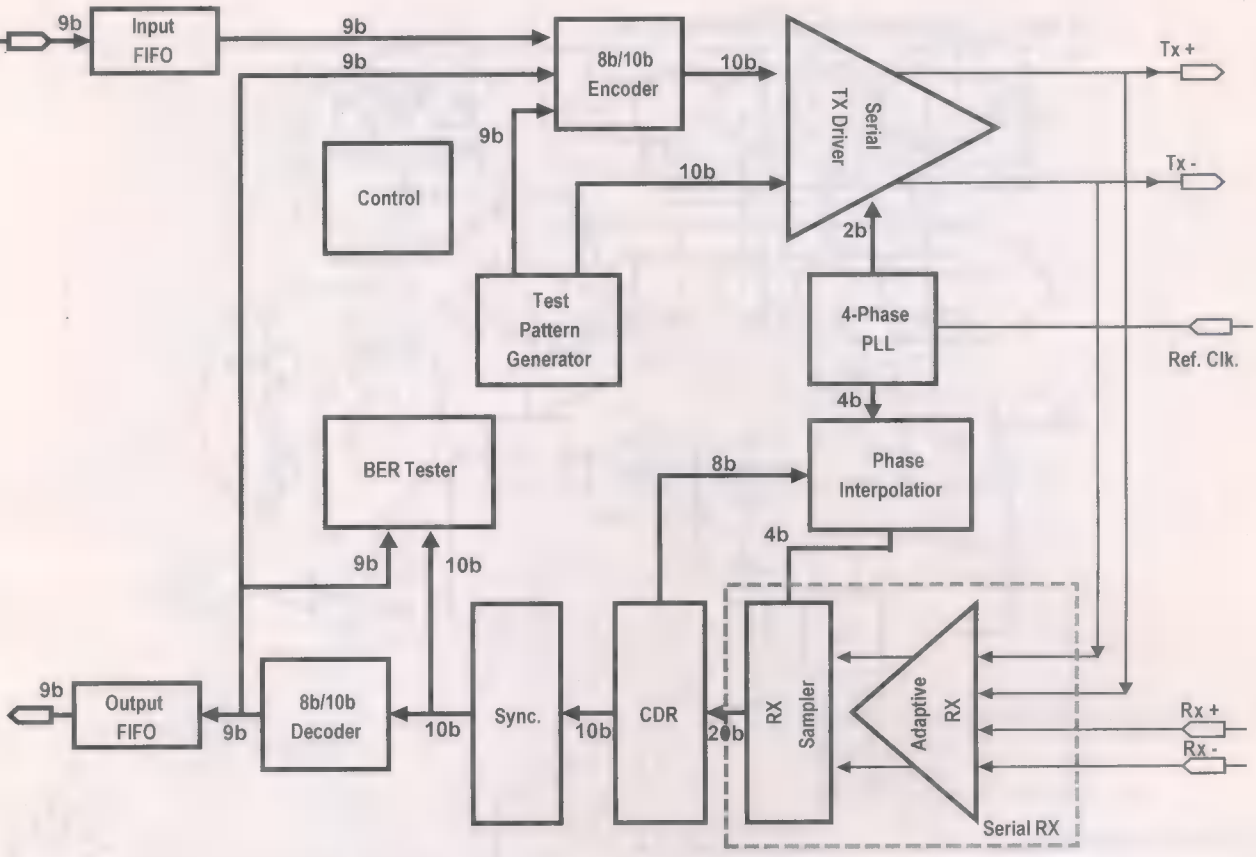


Fig. 1. Block diagram of the transceiver.

Section III explains circuit implementation of each sub-block. Finally, experimental results are given in Section IV, and conclusions presented in Section V.

II. Chip Architecture

The 10-Gb XAUI specification, from the 10-G Ethernet standard 802.3ae, defines the chip-to-chip interconnect protocol as a 12.5-Gb/s full duplex raw data rate with 3.125 Gb/s per channel on four channels [1]. The implementation described in this paper targets the XAUI specification.

The transceiver uses a four-phase clock with half-rate frequency. This clocking scheme enables two-level of input muxing on the transmit (TX) side, and 2X oversampling on the receive (RX) side. The binary CDR uses the 2X oversampled data to recover the clock using a phase-tracking method. A mixed-mode adaptive equalizer is used to reduce ISI. Fig. 1 shows a simplified block diagram of the transceiver.

In the transmit path, input FIFO performs rate matching between 10-Gb Media Independent Interface (XGMII) and XAUI. An 8 b/10 b encoder converts the input octet data with a control bit to a 10-bit coded word. This encoder limits the maximum run length less to than 5 and as a result, every symbol has timing information. Furthermore, it guarantees dc balance because the coded word has balanced 1s and 0s.

The serial TX driver then serializes these coded words, or a test pattern generated from a test pattern generator. It uses input multiplexing (muxing) rather than conventional output muxing because an input-multiplexed transmitter has the advantages of small chip area, low power, and low jitter [18]. Two kinds of test patterns are used. One is bit pattern that includes high-frequency, low-frequency, and mixed-frequency pattern. The other is packet pattern specified in 802.3ae that consists of continuous jitter and continuous random jitter.

In the receive path, a mixed-mode adaptive equalizer reduces the ISI induced from the channel to slim the pulses and make the "eye" open. The two-tap adaptive equalizer consists of a 1-bit delay cell, preamp, TX modeler and tap adaptation circuitry. The 1-bit delay cell delays the analog input by one unit interval (UI) using a delay cell. The delay amount is controlled by a PLL locked to an external reference clock. Tap adaptation uses the sign-sign least-mean-square (LMS) algorithm due to its simplicity of implementation, and it is implemented in the digital domain.

The sampler sequentially latches the output of an adaptive equalizer using the four-phase PLL clocks and generates 2X oversampled data. The CDR circuit extracts the timing information from the 2X oversampled data and feeds the correct sampling timing to the RX sampler using the phase interpolator. The phase interpolator mixes two clock signals selected by the CDR circuit and generates an interpolated clock.

Finally, the synchronizer finds the word boundary from the bit stream using a comma detector and an 8 b/10 b decoder recovers the transmitted octet from the coded words. Output FIFO offers the capabilities of rate matching and channel alignment among four channels using ordered sets for channel alignment $||A||$ as specified in 802.3ae.

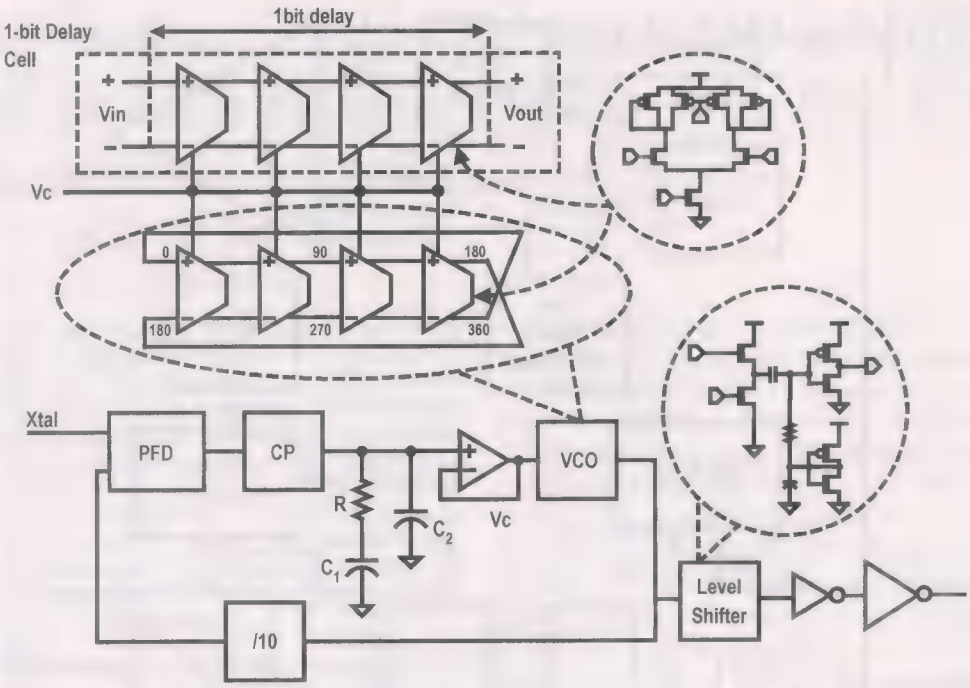


Fig. 2. Block diagram of PLL and 1-bit analog delay cell.

III. Circuit Implementation

A. Clocking and Signaling

This transceiver contains an on-chip clock generation PLL to provide global four-phase half-rate clocks. Since the jitter performance of the PLL ultimately determines the transceiver performance, the clock generation PLL is one of the most important parts of transceiver. In order to achieve low-jitter operation, a PLL design requires buffer stage designs with low supply and substrate noise sensitivity. For robustness, this transceiver employs a self-biased PLL that provides a very broad frequency range, minimized supply and substrate noise induced jitter, and a high input tracking bandwidth [12]. The intrinsic immunity to process technology and environmental variability of self-biasing also gives more stability to the PLL. Fig. 2 depicts the block diagram of this PLL.

Deterministic jitter usually comes from the phase mismatch of the PLL. To meet the jitter requirements at the near end, phase mismatch should be less than 15.3° (0.85 UI). Careful layout was used to avoid mismatches among delay cells and clock signal paths. In order to reduce the noise coupling from the substrate, fully differential design and decoupling capacitors were used. To isolate the PLL from the noisy transmitter and digital circuitry, guard rings and separated power pins were used also.

The differential buffer delay stage used in the PLL requires an inverter chain to supply clocks at rail-to-rail level. This high-frequency level shifter has a bandpass type transfer function and reduces low-frequency noise caused by the source follower and other circuits. Since the inverter with input and output shorted has geometry scaled proportional to the inverters in the inverter chain, it gives an optimal input dc bias level.

A 1-bit delay cell shown in Fig. 2 gets control voltage from the PLL and yields exact 1-bit time T, whenever the PLL is in locked state. These cascaded delay stages can be used as an analog delay cell which is one of the key components of a continuous time analog equalizer, and will be described later in the adaptive equalizer section.

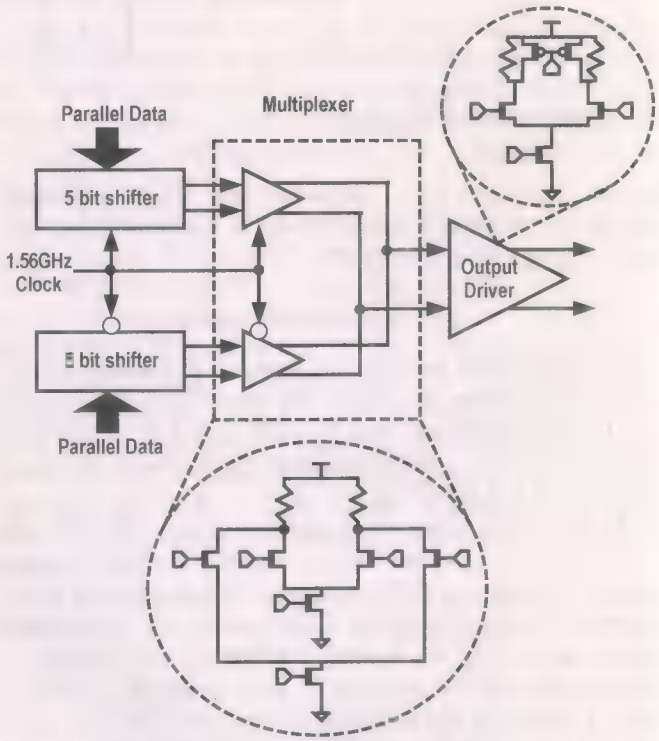


Fig. 3. Block diagram of input-multiplexed transmitter using shifters.

In general, since input-multiplexed transmitters require smaller layout area and have smaller parasitic components at the output node, they achieve better performance than output-multiplexed transmitters [17]. This transceiver also

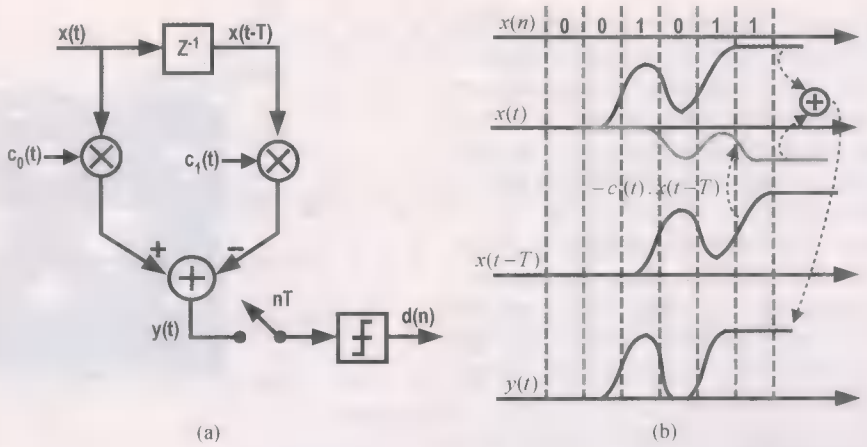


Fig. 4. (a) A simplified block diagram of the continuous-time forward equalizer and (b) operation of the equalizer in time domain.

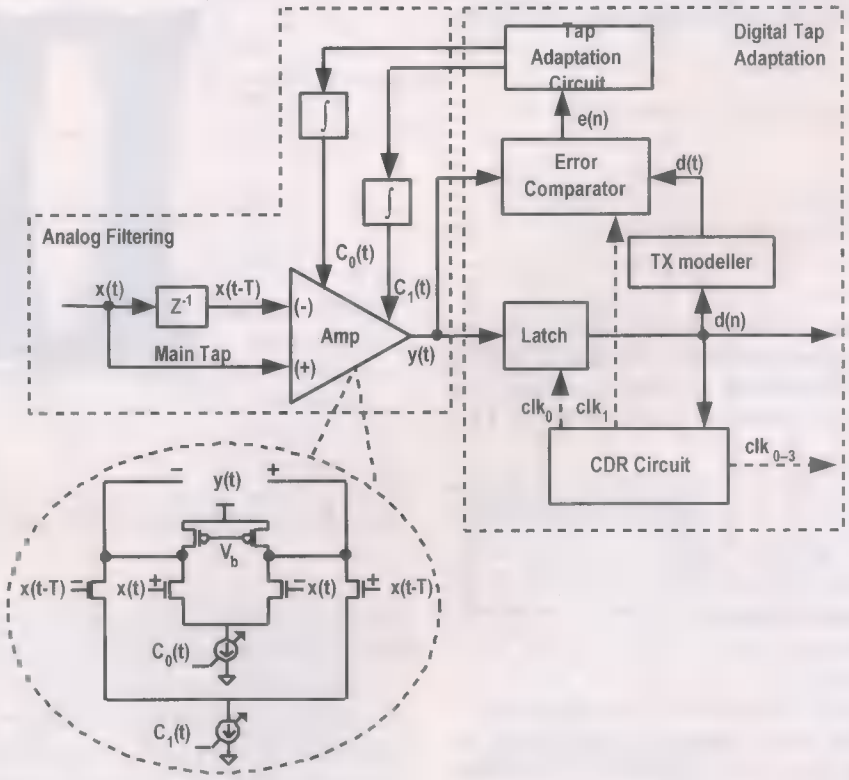


Fig. 5. Block diagram of adaptive equalizer.

adopts an input-multiplexed transmitter using shifters. Fig. 3 shows the transmitter that comprises two 5-bit shifters, a multiplexer and an output driver. The shifters load 10-bit data at every fifth rising edge of the 1.56-GHz clock. One shifter transfers data to the multiplexer at every rising edge and the other at every falling edge of the 1.56-GHz clock. The multiplexer serializes two outputs of the shifters and the output driver transmits 3.125-Gb/s data through the channel.

B. Adaptive Equalizer

From a time-domain viewpoint, channel attenuation forces transferred symbols to spread in time and to interfere each other (ISI). The equalizer in the receiver side sharpens the transition edges of the signal. Sharper transition edges result in wide data eye openings and larger timing margin for signal detection. This effect mainly appears at the most high-frequency bit sequence, repeated "01" as shown Fig. 4(b). The figure illustrates the operation of analog filtering in the time domain.

Fig. 5 illustrates a mixed-mode adaptive equalizer with a two-tap LMS adaptation loop. The analog filtering part realizes an analog transversal equalizer (ATE) and performs high-speed filtering, while the digital tap adaptation part updates the coefficients based on the decision result. The analog filtering equation is

$$y(t) = c_0(t) \cdot x(t) + c_1(t) \cdot x(t - T) \tag{1}$$

where T is the symbol period.

The analog circuit comprises a variable gain amplifier, an analog delay line, a transmitter modeler and an error com-

parator. The variable gain amplifier performs the main filtering functions. It uses two differential pairs with connected output to the same PMOS loads. Tail current sources act as gain modifiers for each tap and their control voltages are the tap coefficients. Each differential pair multiplies the tap coefficients by its input and the resulting currents are summed at the PMOS load to yield the estimated signal $y(t)$.

The transmitter modeler generates the reference signal $d(t)$ according to the digital value extracted from the estimated signal $y(t)$. The error comparator compares the generated ideal transmit waveform of the transmitter modeler with the estimated signal. The compared result is sampled and fed to the digital tap adaptation circuit to update the tap coefficients.

The coefficients $c_k(t)$ in the equalizer can be adapted using the sign-sign LMS algorithm [7]. Charge pumps are used to update the analog coefficient from the output of digital tap adaptation circuitry. The update equations for the equalizer coefficients are

$$C_k(n+1) = C_k(n) + \mu \cdot \text{sign}\,[e(n)] \cdot \text{sign}\,[x(n-kT)] \quad (2)$$

where

$$\text{sign}\,[e(k)] = \text{sign}\,[(y(t) - d(t))\,|_{t=kT}] \quad (3)$$

and μ is scaling factor.

An important advantage of using the sign-sign LMS algorithm is the simplicity of implementation for the multiplication operation in (2). Some potential problems with analog filters such as offset and gain errors are mitigated by the LMS algorithm [18].

This adaptive equalizer employs cascaded differential buffer delay stages to realize a 1-bit delay T. If the PLL locks to an external clock reference, the resulting VCO control voltage makes delay of four-delay cell 180° phase shift as shown in Fig. 2, because an N-stage oscillator generates one cycle of oscillation after propagating through each stage two times. The VCO control voltage feeds the analog 1-bit delay cell also and the cascaded delay stages then yield a delay of half an oscillation cycle, that is, a precise 1-bit delay time because half-rate clocking has been used. Therefore the analog delay line always generates a 1-bit delay time automatically whenever the PLL is in locking state.

A generated $2^{16}-1$ pseudo-random bit sequence (PRBS) at 3.125-Gb/s signal at the end of a 50-cm PCB trace was supplied to the equalizer with a proper setting. Fig. 6 illustrates the result of the HSPICE simulation. As may be seen, the eye is completely open with sufficient margin for the demultiplexing sampler.

C. Delay-Immune Clock Data Recovery (CDR)

In addition to adaptive equalization, the CDR circuit, which retrieves the clock from the nonreturn-to-zero (NRZ) data, is one of the key components of the receiver. It extracts the clock information from the data transitions and adjusts the phase of the sampling clocks. In the tracking phase detection technique, traditional proportional tracking data PLLs offer good loop stability and bandwidth, but generally suffer from a systematic

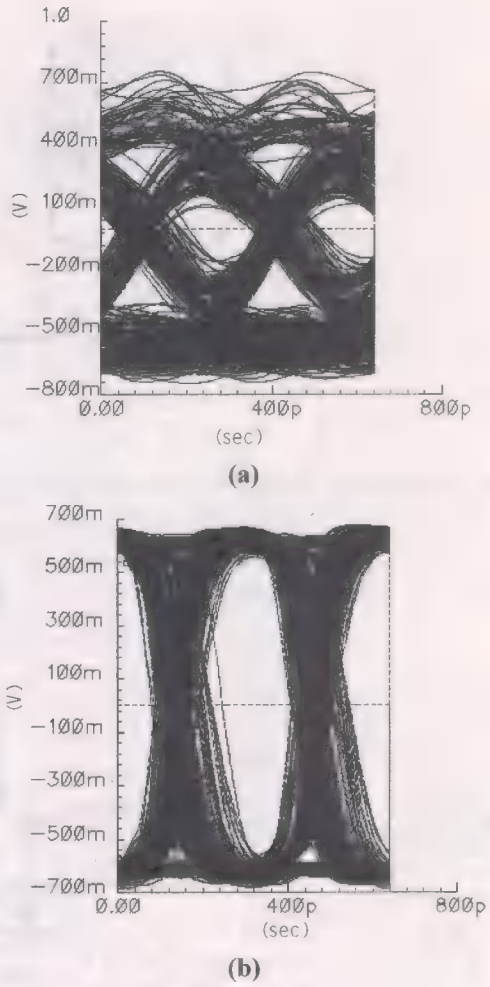


Fig. 6. Simulated eye diagram of adaptive equalizer (a) input and (b) output.

phase offset and long lock time. The binary CDR technique potentially provides a higher tracking bandwidth and greater robustness to phase noise than the PLL based algorithm, but the jitter performance is limited by the resolution [19]. Specifically, the jitter of a binary CDR is set by the minimum resolution of the phase interpolator because of its bang-bang operation [6]. If there are delays in the recovery loop, however, the jitter will be more than the minimum resolution. In this paper, we introduce a novel CDR algorithm that has immunity to the effect of delays in recovery loop.

Fig. 7 shows examples of various CDR algorithms in operation. It is assumed that the incoming data stream has some frequency offset from the reference clock, as allowed by IEEE 802.3ae standard. In the case that a CDR circuit has no delay, Fig. 7(a) immediately updates the timing, and its recovered clock jitter is limited by one minimum resolution of the phase interpolator. In the case of a CDR circuit with delays in the recovery loop, however, it has more jitter due to the delayed timing update, as shown in Fig. 7(b).

To eliminate the effect of delays in the recovery loop, this transceiver adopts a delay-immune CDR algorithm. The motivation for developing a delay-immune CDR algorithm is that the CDR circuit should ignore the excess UP/DN indication caused

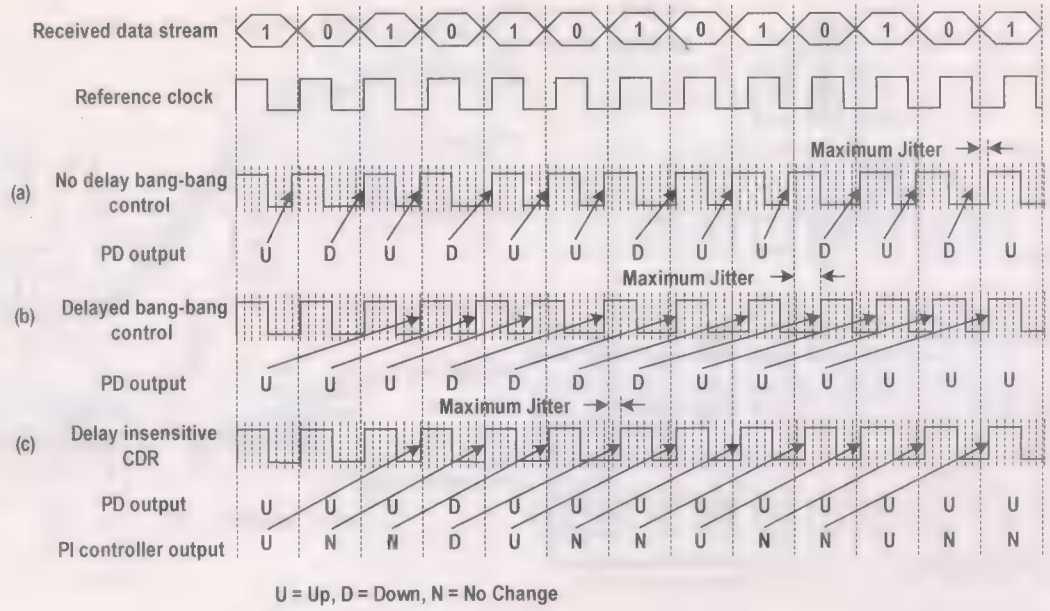


Fig. 7. Examples of various CDR algorithms operating, showing maximum jitter.

from the delays in the recovery loop. To ignore the false indication, the CDR circuit compares the current UP/DN value to the previous UP/DN values of the same number as the delays in the recovery loop. If the current UP/DN value is same as the previous UP/DN values, the CDR circuit does not change the current timing, since the current UP/DN value is generated before the timing updates of the previous UP/DN values due to the delays in the recovery loop. The accumulator in the recovery loop does this operation. As a result, the CDR circuit achieves ideal bang-bang operation and the recovered clock jitter is limited by one minimum resolution of the phase interpolator, as shown in Fig. 7(c).

Fig. 8 shows an implementation block diagram of the delay-immune CDR circuit. Clock recovery employing a dual-loop phase-selection and phase-interpolation scheme [19] is used. A multi-phase PLL supplies evenly spaced phases and clock recovery preformed by the phase-selection and phase-interpolation loop is completely independent of the PLL. A multiplexer selects a pair of adjacent clock phases to define a phase interval for interpolation. The phase interpolation is then supplies sampling clock to the input samplers. The input samplers sample the output of the adaptive equalizer by 2X oversampling to yield center samples and transition samples. These samples are aligned to give Din[9:0] and Dt[9:0] respectively. The transition detector generates the Up[9:0] and Down[9:0] vectors from the input Din[9:0] and Dt[9:0] vectors. The 8 b/10 b encoder ensures there is at least one transition in every coded word, and the comparator counts the number of 1s in each vector and compares their values. As an output, it generates an Inc/Dec signal according to the compared result. A control block is used to a prevent phase discontinuity at quadrant crossings [19]. The accumulator detects a false indication caused by the delays in the loop, and the final phase selection state is latched and fed to the multiplexer and the phase interpolator.

The comparator implementation is straightforward and comprises a binary 10-bit adder to encode bit vector to binary

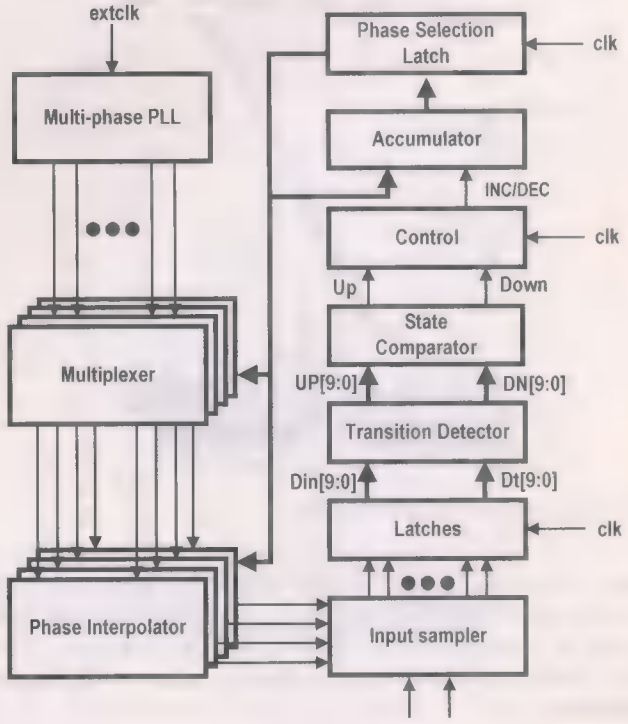


Fig. 8. Clock and data recovery architecture.

number and a 4-bit binary comparator. However, it is difficult to meet the timing requirements using this straightforward digital implementation. Fig. 9 shows a novel approach to perform the same function. The basic idea of this algorithm is trellis passing according to each bit value of Up[9:0] and Dn[9:0]. Instead of counting the number of 1s in the vector, the state is changed for each bit value. If Up[n] and Dn[n] have the same value, the next state has the same position. If they differ, however, the state moves toward the direction of 1. Because all the 10 bits apply at the same time, the total time delay is 10 times that of one state transition. Fig. 9 shows an implementation example of the

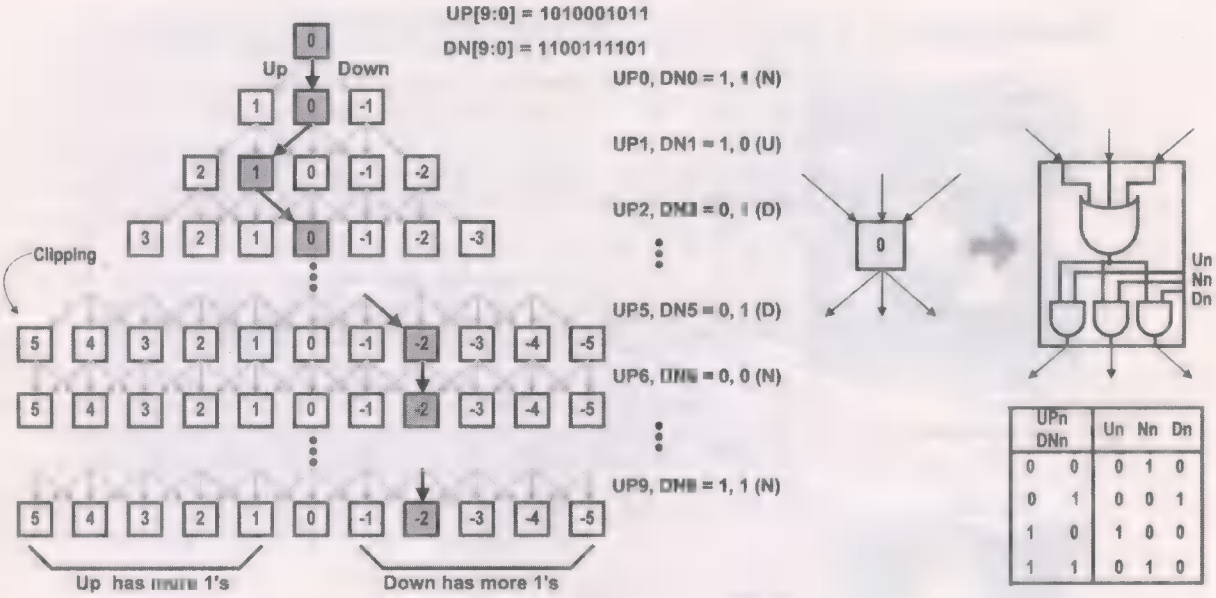


Fig. 9. State comparator algorithm and implementation.

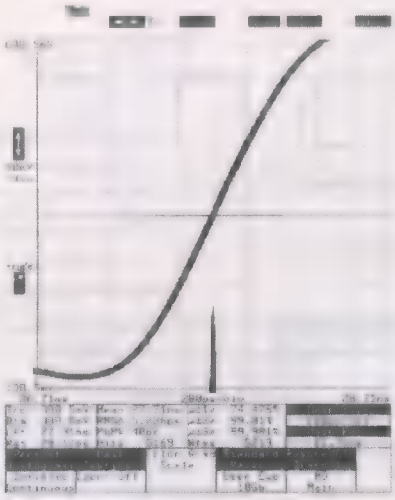


Fig. 10. Performance of self-biased PLL.

state comparator. It only consists of three two-input AND gates and one three-input OR gate, and by using Boolean operation, it can be converted to NAND–NAND or NOR–NOR logic. It has very simple implementation architecture allowing high-speed operation.

IV. Measurements

This transceiver chip was fabricated in 0.18-μm standard CMOS technology with 1.8-V supply voltage and packaged in 256-pin PBGA. It provides 12.5-Gb/s full duplex raw data rate for a single 10-Gb XAUI. The power consumption is 178 mW per channel and total 718 mW at 3.125-Gb/s full duplex (Tx/Rx simultaneous) data rate.

Fig. 10 shows the performance of the PLL. It locks to 156.25-MHz crystal oscillator reference and gives 5.036-ps (rms) jitter and 40-ps (p-p) jitter. The PLL rms jitter reduces about 1.8%$-\Delta$rms jitter/%$-\Delta V_{dd}$ as supply voltage increases. The input multiplexing transmitter performance is shown in

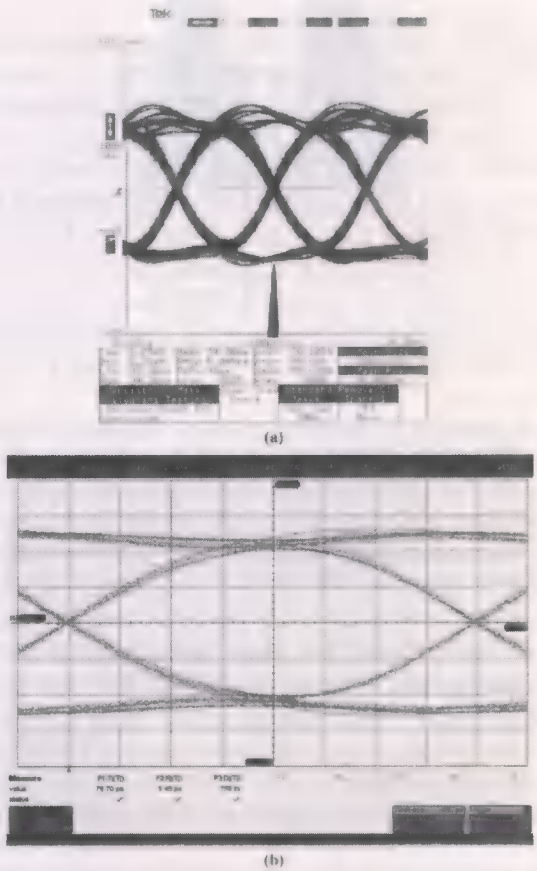


Fig. 11. The TX eye diagram and performance (a) in terms of RMS jitter and (b) in terms of total jitter.

Fig. 11. Fig. 11(a) is a result with digital sampling oscilloscope showing 5.045-ps (rms) jitter and 46-ps (p-p) jitter. Fig. 11(b) is a result with LeCroy SDA6000 equipment showing 78.7-ps total jitter, 5.46-ps random jitter, and 756-fs deterministic jitter. The transmitter output has differentially adjustable amplitude with a maximum of 1600 mV from 800 mV.

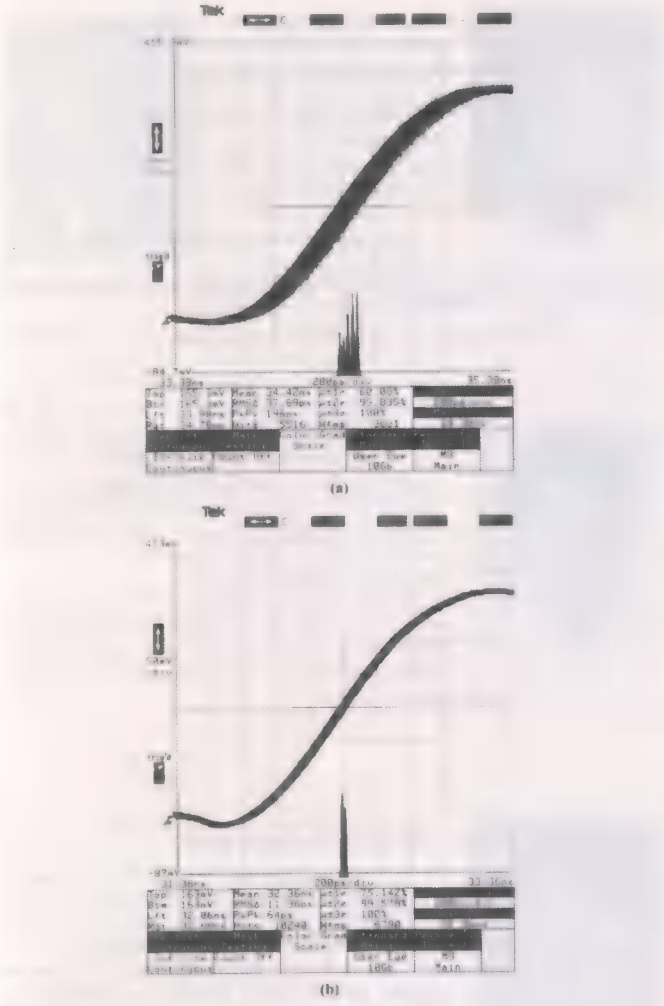


Fig. 12. Performance comparison of CDR algorithm. (a) Conventional bang-bang algorithm and (b) delay-immune algorithm.

Fig. 13. Transceiver die photograph.

Fig. 12 shows the performance of delay-immune CDR circuit. Fig. 12(a) shows bang-bang controlled CDR performance with delays and has a jitter of 37.69 ps (rms) and 196 ps (p-p), while Fig. 12(b) shows the delay-immune CDR jitter performance of 11.36 ps (rms) and 64 ps (p-p).

TABLE I
PERFORMANCE SUMMARY

Transmitter performance	
Maximum transmitter rate	13.2 Gb/s @1.8V, 4 ch
Output Jitter @3.125 Gb/s (rms)	46 ps (p-p), 5.045 ps (rms)
Output jitter @3.125 Gb/s (total jitter)	78.7 ps (TJ), 5.46 ps (RJ), 0.756 ps (DJ)
Swing level	800 mV ~ 1.6 V (Diff.)
Receiver performance	
Maximum receive rate	13.2 Gb/s @1.8 V, 4ch
Recovered Clock jitter @3.125 Gb/s	64 ps (p-p), 11.36 ps (rms)
BER	4.5×10^{-15}
Power dissipation @3.125Gb/s, 1.8V	718 mW (4ch) / 178.2 mW (1ch)
PLL	75.6mW (1ch)
Rx + CDR	27mW (1ch)
Tx	46.8mW (1ch)
Digital (Tx/ Rx)	28.8mW (14.4mW/ 14.4mW) (1ch)
Area	2.3 mm x 2.3 mm (core)

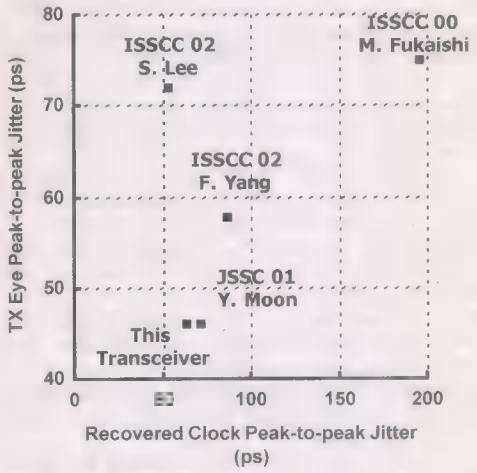


Fig. 14. Performance comparison table for previous serial-link transceivers.

The BER measurements are performed using a built-in pattern generator and BER tester. With various bit patterns and packet patterns specified in 802.11ae, the transceiver shows the BER performance lower than $< 4.5 \times 10^{-15}$.

All the building blocks of the multiphase PLL are fully integrated on the chip including the loop filter. The chip occupies 2.3 mm × 2.3 mm of die area. The transceiver die photo is shown in Fig. 13. Table I summarizes the transceiver chip performance and Fig. 14 shows a comparison matrix with previous work.

V. Conclusion

A four-channel 3.125-Gb/s/ch CMOS serial-link transceiver is fabricated in a 0.18-μm CMOS process. An input multiplexing transmitter with a low-jitter PLL shows only 46-ps peak-to-peak jitter. For a receiver, a mixed-mode LMS adaptive equalizer is implemented to reduce ISI and to improve BER performance. A delay-immune CDR algorithm is proposed and implemented for clock recovery loop stability. Recovered clock jitter is measured to 64 ps (p-p). Because of these techniques, the measured BER performance of the overall transceiver is lower than $< 4.5 \times 10^{-15}$.

References

[1] *IEEE 802.3ae Standard Draft 4.1*, Feb. 2002.

[2] R. Farjad-Rad *et al.*, "A 0.4-μm CMOS 10-Gb/s 4-PAM pre-emphasis serial link transmitter," *IEEE J. Solid-State Circuits*, vol. 34, no. 5, pp. 580–585, May 1999.

[3] ——, "A 0.3 μm CMOS 8-Gb/s 4-PAM serial link transceiver," *IEEE J. Solid-State Circuits*, vol. 35, no. 5, pp. 757–764, May 2000.

[4] K. Azadet *et al.*, "Equalization and FEC techniques for optical transceivers," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 317–327, Mar. 2002.

[5] R. Alini *et al.*, "A 200-Msample/s Trellis-coded PRML read/write cannel with analog adaptive equalizer and digital Servo," *IEEE J. Solid-State Circuits*, vol. 32, no. 11, pp. 1824–1838, Nov. 1997.

[6] M. Fukaishi *et al.*, "A 20-Gb/s CMOS multichannel transmitter and receiver chip set for ultra-high-resolution digital displays," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1611–1618, Nov. 2000.

[7] J. Bucklew, D. Johns, and W. Snelgrove, "Comparison of DC offset effects in four LMS adaptive algorithms," *IEEE Trans. Circuits Syst. II*, vol. 42, no. 3, pp. 176–185, Mar. 1993.

[8] V. Minuhin and V. Kovner, "Adaptive, analog, continuous time-domain equalization for sampled channels in digital magnetic recording," in *IEEE Int. Magnetics Conf. Dig.*, Apr. 1997, p. CR-08.

[9] H. Yamamoto and S. Mori, "Performance of a binary quantized all digital phase-locked loop with a new class of sequential filter," *IEEE Trans. Commun.*, vol. COM-26, no. 1, pp. 35–45, Jan. 1978.

[10] H. Ransjin and P. O'Conner, "A PLL-based 2.5 Gb/s GaAs clock and data regenerator IC," *IEEE J. Solid-State Circuits*, vol. 26, no. 10, pp. 1345–1353, Oct. 1991.

[11] R. Walker, C. Stout, and C.-S. Yen, "A 2.488 Gb/s Si-bipolar clock and data recovery IC with robust loss of signal detection," in *IEEE ISSCC Dig. Tech. Papers*, 1997, pp. 246–247.

[12] J. G. Maneatis, "Low-jitter process-independent DLL and PLL based on self-biased techniques," *IEEE J. Solid-State Circuits*, vol. 31, no. 11, pp. 1723–1732, Nov. 1996.

[13] M. Lee *et al.*, "A 90 mW 4 Gb/s equalized I/O circuit with input offset cancellation," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2000, pp. 252–253.

[14] P. Larsson, "An offset-cancelled CMOS clock-recovery/demux with a half-rate linear phase detector for 2.5 Gb/s optical communication," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2001, pp. 74–75.

[15] S. B. Anand and B. Razavi, "A CMOS clock recovery circuit for 2.5 Gb/s NRZ data," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 432–439, Mar. 2001.

[16] S.-H. Lee *et al.*, "A 5-Gb/s 0.25-um CMOS jitter-tolerant variable-interval oversampling clock/data recovery circuit," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1822–1830, Dec. 2002.

[17] M.-J. E. Loo *et al.*, "A 90 mW 4 Gb/s equalized I/O circuit with input offset cancellation," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2000, pp. 252–253.

[18] J. E. C. Brown *et al.*, "A CMOS adaptive continuous-time forward equalizer, LPF, and RAM-DFE for magnetic recording," *IEEE J. Solid-State Circuits*, vol. 34, no. 2, pp. 162–169, Feb. 1999.

[19] F. Yang *et al.*, "A CMOS low-power multiple 2.5–3.125-Gb/s serial link macrocell for high IO bandwidth network ICs," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1813–1821, Dec. 2002.

Jinwook Kim received the B.S. degree in electronics and electrical engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 1996, and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejon, Korea, in 1998. Since 1998, he has been working toward the Ph.D. degree at KAIST, focusing on high-speed serial-link transceiver.

Since 2001, he has been with Berkäna Wireless Korea, Inc., Seoul, Korea, where he is a Senior Design Engineer. His research interests include high-speed serial-link transceiver system design and optimization.

Jeongsik Yang received the B.S. degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1993 and the M.S. and Ph.D. degrees in electrical engineering and computer science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1996 and 2002, respectively.

In 2002, he joined Berkäna Wireless Inc., Campbell, CA, where he is a Senior Design Engineer. His research interests include high-speed serial-link transceiver, RF transceiver, analog IC design, and mixed-mode signal processing IC design.

Sangjin Byun was born in Seoul, Korea, in January 1976. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1997, 1999, and 2004, respectively.

From 2001 to 2004, he was with Berkäna Korea, Inc., Seoul, Korea, where he was involved in designing CMOS analog circuits for Bluetooth and SERDES transceiver ICs. In 2004, he joined the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, where he is a Senior Member of Engineering Staff. His research interests include CMOS analog/mixed-mode circuit design and layout techniques for communication ICs.

Hyunduk Jun received the B.S. degree in information and communications engineering from Chungbuk National University, Korea, in 2002.

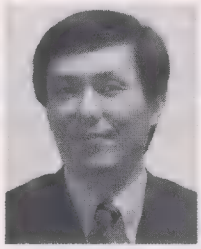
Since February 2003, he has been a Staff Application Engineer with Berkäna Wireless Inc., Seoul, Korea. His current research interests include signal circuit design in wireless communication.

Jeongkyu Park received the B.S. degree in electronics engineering from Pukyong National University, Busan, Korea, in 2002.

From 2001 to 2004, he worked for Berkäna Wireless Korea, Inc., where he is currently a System Engineering Staff Member with the RF baseband system team. His research interest is in embedded systems and microprocessor design.

Cormac S. G. Conroy (S'83–A'85–M'95) received the B.E. degree in electrical engineering from University College, Cork (UCC), Ireland, in 1985, and the M.Eng.Sc. degree from the National Microelectronics Research Centre, UCC, in 1987. In 1994, he received the Ph.D. in electrical engineering from the University of California, Berkeley, for work on high-speed A/D conversion in CMOS with particular focus on pipeline, time-interleaved, and digitally calibrated architectures.

After graduating, he was with IBM Storage Systems Division, San Jose, CA, and in August 1994 he joined DataPath Systems where he worked on highly integrated mixed-signal ICs for storage and communications. At DataPath, he was Program Manager leading the design and development efforts in CMOS ADSL analog front-end ICs over multiple product generations. DataPath was acquired by LSI Logic in May 2000. In early 2001, with Dr. Beomsup Kim, he co-founded Berkäna Wireless Inc., a Silicon Valley-based fabless semiconductor company that develops highly integrated CMOS RF solutions for wireless connectivity.



Beomsup Kim (S'87–M'90–SM'95–F'04) received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1983 and 1985, respectively, and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1990.

From 1990 to 1991, he was with Chips and Technologies, Inc., San Jose, CA, where he was involved in designing high speed-signal processing ICs for disk drive read/write channel. From 1991 to 1993, he was with Philips Research, Palo Alto, CA, and conducted research on digital signal processing for video and wireless communication. In 1994, he joined Korea Advanced Institute of Science and Technology (KAIST), Daejon, Korea, as a faculty member with the Department of Electrical Engineering. During 1999, he took a sabbatical leave and was a visiting scholar at Stanford University, Stanford, CA, and consulted for Marvell Semiconductor Inc., San Jose, CA, on the Gigabit Ethernet (802.11ab) and wireless LAN (802.11b) DSP architecture. In 2001, he co-founded Berkana Wireless Inc. and is now CTO/VP Engineering of the company. His research interests include mixed-mode signal processing IC and system design for wireless communication, telecommunication, disk drive, local area network, high-speed analog IC design, and VLSI system design.

Dr. Kim was a co-recipient of the Best Paper Award (1991) for the IEEE JOURNAL OF SOLID-STATE CIRCUITS and received the Department (EE) Best Lecture Award at KAIST (1997). Between June 1993 and June 1995, he served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING. In 1999, he was one of four lecturers for the Gigabit Ethernet short course at the IEEE International Solid-State Circuit Conference.

Low-Voltage Low-Power LVDS Drivers

Mingdeng Chen, *Member, IEEE*, Jose Silva-Martinez, *Senior Member, IEEE*, Michael Nix, and Moises E. Robinson, *Member, IEEE*

***Abstract*—Two low-voltage low-power LVDS drivers used for high-speed point-to-point links are discussed. While the previously reported LVDS drivers cannot operate with low-voltage supplies, the proposed double current sources (DCS) LVDS driver and the switchable current sources (SCS) LVDS driver are suitable for low-voltage applications. Although static current consumption is greater than the minimum amount required by the signal swing, the DCS LVDS driver is simple and fast. The SCS LVDS driver, by dynamically switching the current sources, draws minimum static current and reduces the power consumption by 60% compared to previously reported realizations. Both drivers were fabricated in a standard 0.35-μm CMOS process; they are compliant with LVDS standards and can operate at data rates up to gigabits-per-second.**

***Index Terms*—Back-plane drivers, fast data communication circuits, input/output (I/O) drivers, low-voltage differential signaling (LVDS), low-voltage low-power integrated circuits.**

I. Introduction

THE ever-increasing processing speed of microprocessor motherboards, optical transmission links, chip-to-chip communications, etc., is pushing the off-chip data rate into the gigabits-per-second range. While scaled CMOS technologies continue to enhance on-chip operating speeds, off-chip data rates have gained little benefit from the increased silicon integration. This is primarily due to the excessive power consumption necessary for driving impedance-controlled electrical interconnects, which leads to an increase in costs related to packaging and thermal management [1]. In the past, off-chip high data rates were achieved by massive parallelism, with the disadvantages of increased complexity and cost for the IC package and the printed circuit board (PCB). Therefore, it is beneficial to move the off-chip data rate to the range of Gb/s-per-pin or above. Reducing the power consumption is also critical for battery-powered portable systems as well as some other systems in order to extend the battery life and reduce the costs related to packaging and additional cooling systems.

Scalable Coherent Interface (SCI) is a high-speed packet transmission protocol that efficiently provides the functionality of bus-like transactions (read, write, lock, etc.), but it uses a collection of fast point-to-point links instead of physical buses to reach higher speeds. The initial physical implementations were based on emitter coupled logic (ECL) signal levels [2], which consume more power than is practical in a low-cost workstation environment. Low-voltage differential signaling (LVDS) is a technology developed to provide a low-power and low-voltage alternative [3] to ECL and other high-speed I/O interfaces for point-to-point transmissions. LVDS achieves higher speed and significant power savings by means of a differential scheme for transmission and termination, in conjunction with low voltage swing.

Manuscript received March 2, 2004; revised August 24, 2004.

M. Chen and J. Silva-Martinez are with Texas A&M University, Analog and Mixed-Signal Center, College Station, TX 77843-3128 USA (e-mail: jsilva@ee.tamu.edu).

M. Nix and M. E. Robinson are with Xilinx Inc., Communication Technology Division, Austin, TX 78746 USA.

Digital Object Identifier 10.1109/JSSC.2004.840955

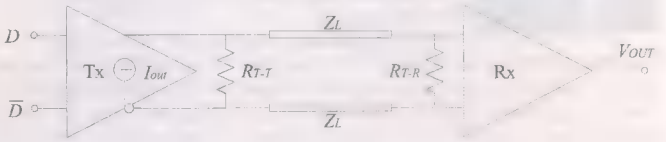


Fig. 1. LVDS interface with termination at the receiver and source ends for gigabits-per-second operation.

In this paper, two low-voltage, low-power, and high-speed LVDS drivers are discussed. Both drivers can operate with data rates of 1 Gb/s and above, and they are fully compatible with IEEE Std 1596.3-1996 [3] for general-purpose links and IEEE Draft P802.3ae/D5.0 [4] for XSBI interfaces. Section II discusses the LVDS interfaces, the typical LVDS drivers, and the design challenges for low-voltage operation. In Section III, the low-voltage, low-power LVDS drivers are discussed and some of the simulation results are also presented. The experimental results and conclusions are addressed in the last two sections.

II. Typical LVDS Drivers

An LVDS interface, as shown in Fig. 1, has a low-voltage swing (250–400 mV); it is connected point-to-point and achieves very high data rates (up to 500 Mb/s per signal pair) and reduced power dissipation [3]. LVDS uses differential data transmission and the transmitter is configured as a switched-polarity current generator. A differential load resistor at the receiver end provides optimum line impedance matching.

Due to the imperfect termination, package parasitics, component tolerances or crosstalk [5], there are reflected waveforms returning to the driver. As data rates push significantly above 500 Mb/s and connectors are added, an additional termination resistor is usually placed at the source end to suppress reflected waves, and the LVDS signaling can be substantially enhanced. Low voltage differential signaling is a standardized data transmission format that is widely used for serial data transmissions; as shown in Fig. 2, a differential signal is centered at a common-mode voltage of about 1.25 V. The maximum magnitude of the differential signal is 400 mV. Typically, the LVDS signal varies in magnitude from 1.05 to 1.45 V.

A typical bridged-switches LVDS driver behaves as a current source with switched polarity as shown in Fig. 3(a) [3]. The bias current I_b is switched through the termination resistors according to the data input, and thus produces the correct

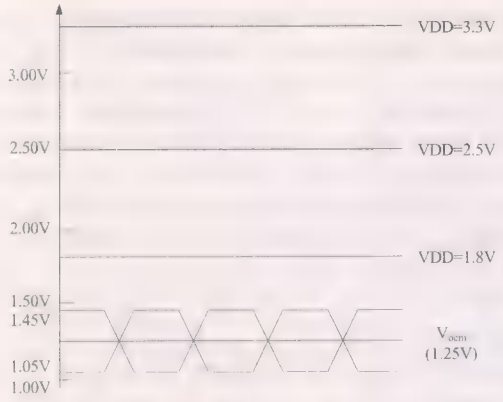


Fig. 2. LVDS signal formatting.

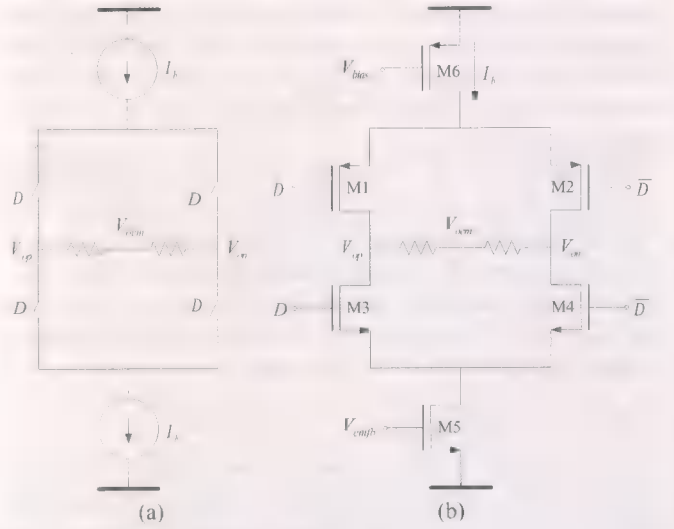


Fig. 3. Typical LVDS driver: (a) macromodel and (b) transistor implementation [3].

differential output signal swing. A possible implementation of the typical LVDS driver is shown in Fig. 3(b). It uses four MOS switches (M1–M4) in a bridged configuration. If switches M1 and M4 are on (D = LOW), the polarity of the output current is positive together with the differential output voltage. On the contrary, if switches M1 and M4 are off (switches M2 and M3 are on), the polarity of the output current and voltage is reversed.

The typical LVDS driver works well if the supply voltage (V_{DD}) is 2.5 V or greater. It is simple and only needs minimum static current consumption to produce the required output signal swing. But when the supply voltage drops below 2 V (e.g., 1.8 V for 0.18-μm CMOS technology), the typical LVDS driver does not have enough headroom in the V_{DD} direction. This is mainly due to the finite on-resistance of the PMOS transistor switches and the large amount of current (nominally 6.4 mA for a signal swing of 320 mV and a 50-Ω termination resistance) flowing through the switches. The voltage drop across the transistor consumes headroom and it demands relatively high voltage supplies for the LVDS driver to operate properly.

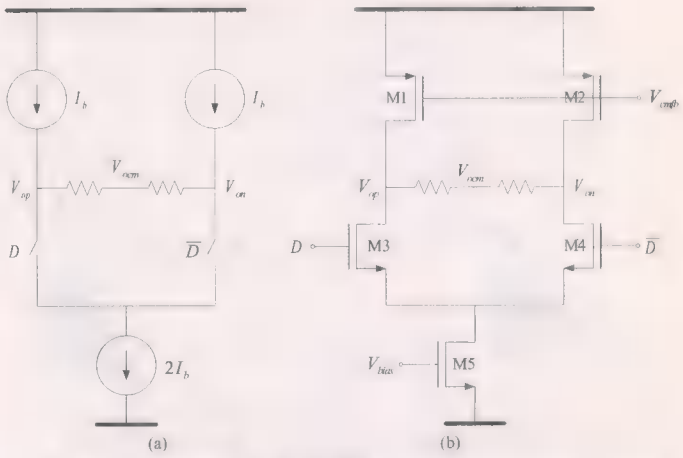


Fig. 4. DCS LVDS driver. (a) Model and (b) potential transistor level realization.

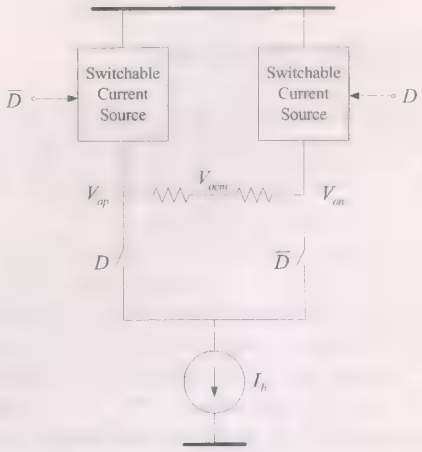


Fig. 5. SCS LVDS driver model.

III. Low-Voltage, Low-Power LVDS Drivers

A. Double Current Sources (DCS) LVDS Driver

A solution to the headroom issue discussed in Section II is to remove the top PMOS switches in the typical LVDS driver [Fig. 3(b)] and replace them by two PMOS current sources, as shown in Fig. 4(a); We call this structure a double current sources (DCS) LVDS driver. In order to produce the same signal swing, the bottom NMOS current source is required to sink $2I_b$, which doubles the static current consumption as required by the output signal swing. Accordingly, the embodiment of Fig. 4(b) consumes more current than the embodiment of Fig. 3(b). In addition, the NMOS transistor switches and the bottom NMOS current source are required to be larger than the corresponding transistors in Fig. 3(b). If an integrated circuit includes a plurality of LVDS drivers, the increased current consumption and transistor dimensions may limit their applications. Also, larger transistor dimensions increase the total pad capacitance and so reduce the pin bandwidth.

B. Switchable Current Sources (SCS) LVDS Driver

Another solution to the headroom issue is shown in Fig. 5. Instead of using two constant current sources at the top, two switchable current sources are used [6]. Depending on the data input, one of the two switchable current sources will

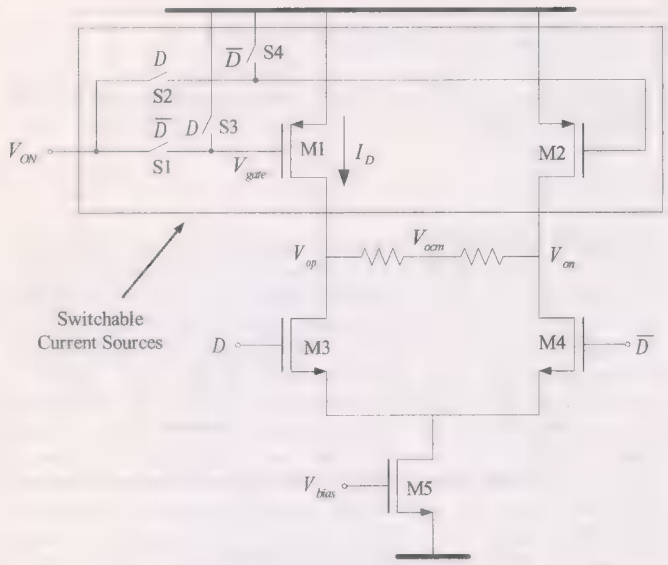


Fig. 6. SCS LVDS driver with control circuit.

conduct current. This current flows through the termination resistors and produces the output voltage swing. Notice that the bottom NMOS current source only needs to sink I_b, leading to minimum static current consumption.

Fig. 6 shows the basic principle behind the proposed SCS LVDS driver. When $V_{\rm ON}$, a reference voltage, is applied to the gate of M1(M2), the transistor conducts a current I_D, which is a copy of a well-controlled reference current, regardless of the process, voltage, and temperature (PVT) variations. Here, transistors M1 and M2 and switches S1, S2, S3, and S4 act as switchable current sources. For instance, when D is LOW (M1 is ON) then M1 conducts current I_D, and it flows throughout the load resistors and M4 to produce the proper output voltage swing.

There are two design issues that need to be addressed for the SCS LVDS driver to operate properly. First, we must determine how to generate the reference voltage $V_{\rm ON}$ such that I_D remains at the proper value regardless of the PVT variations. Second, since the PMOS switchable current sources need to conduct large currents, their transistor dimensions are large as well as their parasitic capacitances. So the question is either how to switch the gate voltages of M1 and M2, or how to quickly charge and discharge the parasitic capacitors at the gates of M1 and M2. The design issues mentioned above are addressed in the SCS LVDS driver shown in Fig. 7; its operation is explained as follows.

The SCS LVDS driver contains two parts: the switchable current source control module and the core of the LVDS driver. The left part of Fig. 7 is the control module, and it is used to generate $V_{\rm ON}$ such that when it is applied to the gate of M1(M2) its drain current I_D is proportional to $I_{\rm ref}$. The cascode transistor M7 and amplifier Amp form a regulated-gain control (RGC) loop. This RGC loop is used to set M6's drain voltage to $V_{D\_\rm ref}(= 1.41$ V). It is important to make sure that the output common-mode voltage and signal swing are maintained; hence the higher output voltage of $V_{\rm op}(V_{\rm on})$ is fixed, and it is defined by $V_{D\_\rm ref}(= V_{\rm ocm\_ref} + V_{o,\rm swing}/2)$, regardless of the PVT variations. $V_{\rm ocm\_ref}$ is the output common-mode reference voltage, and $V_{o,\rm swing}$ is the required signal swing. For instance, for an output common-mode voltage of 1.25 V and an output signal swing of 320 mV, ideally the higher LVDS output voltage $V_{\rm op}(V_{\rm on})$ should be 1.41 V. By setting the drain voltage of M6 to $V_{D\_\rm ref}$, we have good matching for the current mirror composed of M6 and M1 (M2). Another issue worth mentioning is that the switchable current source control module can be shared by several LVDS drivers, but independent buffers are used for each driver in order to minimize the signal feedthrough.

The right part of Fig. 7 is the core of the SCS LVDS driver. The switchable current sources are used to generate current I_D and they are composed of transistors M1 and M2, buffer-connected amplifier Buf-A, switches S1 and S2, and the pull up/down circuits. The pull up/down circuits are used to quickly change the gate voltages of M1 and M2, i.e., to quickly charge or discharge the parasitic capacitors associated with the node $V_{\rm gate}$. The buffer-connected amplifier Buf-A is used to isolate the DC voltage $V_{\rm ON}$ from the data controlled switches. It also provides "fine adjustment" to the gate voltage of M1(M2) when the switch S1(S2) is closed, while the pull up/down circuit, driven by the input data, provides coarse control. The CMFB is used to set the output common-mode voltage to the desired reference voltage $V_{\rm ocm\_ref}$.

The operation of the switchable current sources is explained as follows. If data D is LOW, then switch S1 is ON and switch S2 is OFF. The M1's gate voltage is pulled down to $V_{\rm ON}$ through the pull up/down circuit during the data transition while M2's gate voltage is pulled up close to V_{DD}. M1 conducts current I_D and M2 is OFF. The current I_D flows through the termination resistors and produces the signal swing.

C. Pull Up/Down Circuits

An active pull up/down circuit is shown in Fig. 8 [7]. In this structure, both pull up and pull down sections produce short periods of current pulses at the data's transition edges. These current pulses are used to charge/discharge the parasitic capacitors and so to pull up/down the switchable current source gate voltages. Some design issues are associated with this active pull up/down circuit. First, the circuit itself consumes huge dynamic power since the several delay cells used and the high data rate. Second, the currents produced by the pull up/down circuit are finite and they limit the speed of the charging/discharging process. Also, since the currents are produced by PMOS and NMOS transistors, respectively, the charge injected into the capacitors may not equal the charge extracted from the capacitors. This difference should be supplied by the "Buffer" as shown in Fig. 7, and this requires a fast circuit implementation that demands more power consumption.

Instead of using an active pull up/down circuit, we propose to use passive capacitors C_{PP} driven by the input data for the SCS LVDS driver; the principle of operation is shown in Fig. 9. The passive pull up/down circuit does not have the drawbacks faced by the active pull up/down circuit mentioned above. The capacitor C_{PP}, driven by the input data D, is used to pull up/down M1(M2) gate voltage with drastically reduced transition time and to provide coarse control over the gate voltage $V_{\rm gate}$. The parasitic capacitor C_P associated with the node $V_{\rm gate}$, and capacitor C_{PP} form a capacitive voltage divider. When D goes

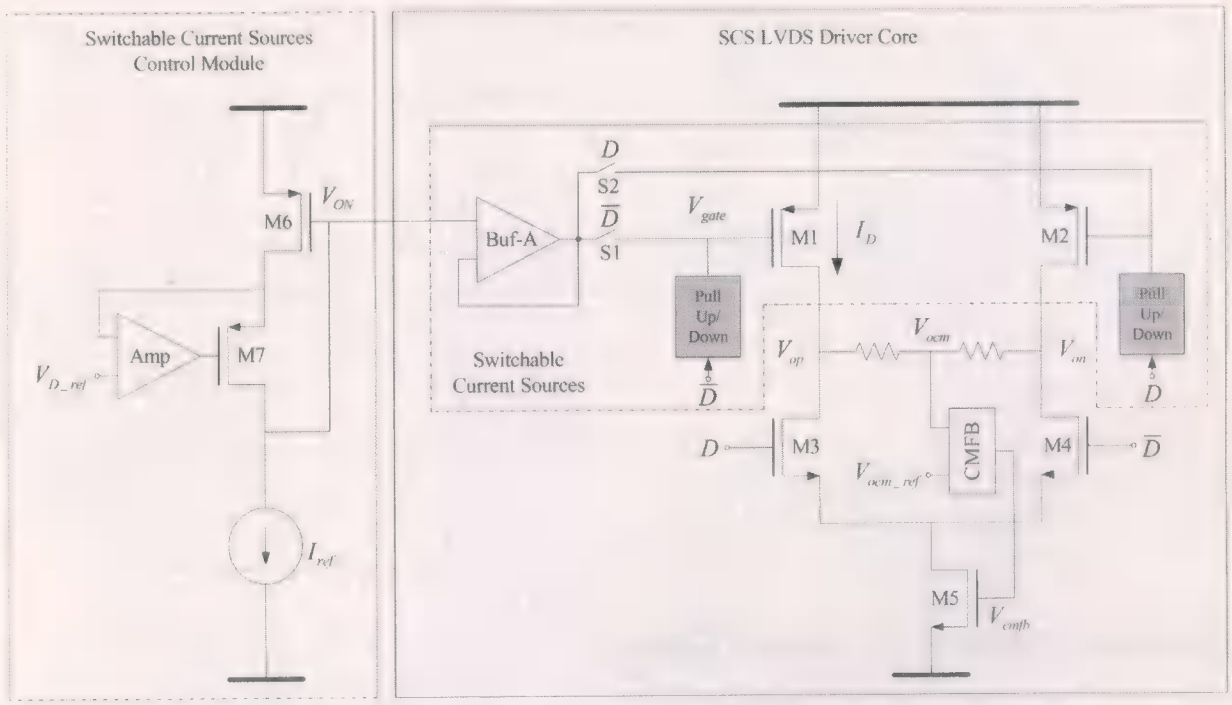


Fig. 7. SCS LVDS driver with active pull up/down circuit auxiliary circuits.

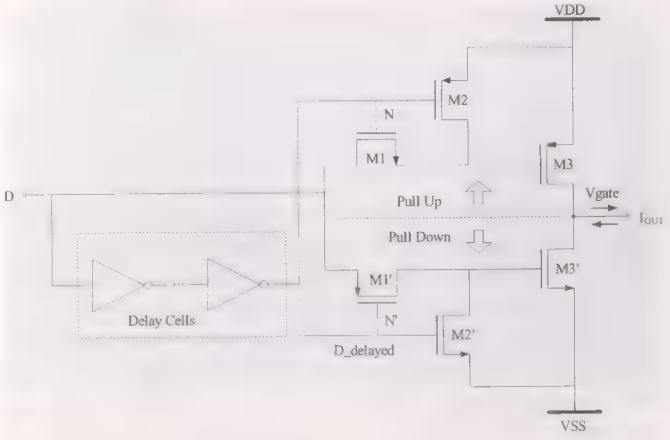


Fig. 8. Active pull up/down circuit [7].

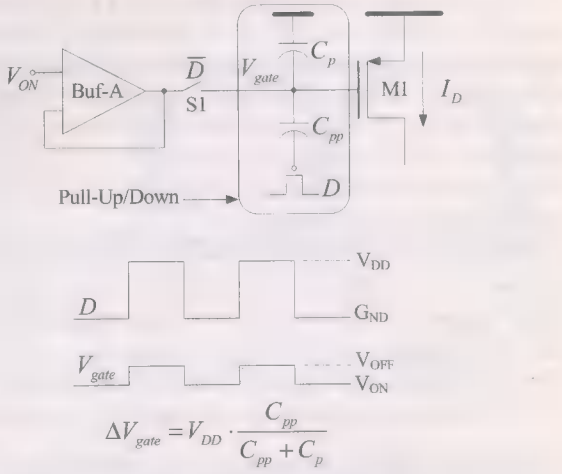


Fig. 9. Passive pull up/down circuit based on charge redistribution.

down, V_{gate} equals V_{ON} and I_D is determined by I_{ref}, while C_{PP} is charged to V_{ON}. During the low-high transition of D, the switch resistance is high and the C_{PP}'s injected charge is mainly absorved by C_P, turning off the transistor. The resulting waveforms of the data and the gate voltage V_{gate} are also shown in Fig. 9. It is easy to show that the M1(M2) gate voltage variation ΔV_{gate} can be expressed as

$$\Delta V_{\text{gate}} = \frac{C_{pp}}{C_{pp} + C_p} \cdot V_{DD} \tag{1}$$

where ΔV_{gate} is defined as $\Delta V_{\text{gate}} = V_{\text{OFF}} - V_{\text{ON}}$. It is assumed that data D varies from V_{DD} to zero.

It is worth mentioning that when the transistor M1 (M2) is turned off, its gate voltage V_{OFF} does not need to be V_{DD}; for fast circuits, it is better for V_{OFF} to be lower than V_{DD} such that the transistor operates in subthreshold region. In this way, we can turn on/off the switchable current sources more quickly and minimize the dynamic power consumption needed to charge/discharge C_{pp} and C_p, as long as the current flowing through the OFF switchable current source I_{OFF} is negligible.

By choosing a proper limit for I_{OFF}, we can find the gate voltage variation ΔV_{gate} such that I_{OFF} does not exceed this limit. Then, the value of the capacitor C_{pp} can be determined as

$$C_{pp} = \frac{C_p \cdot \Delta V_{\text{gate}}}{V_{DD} - \Delta V_{\text{gate}}}. \tag{2}$$

For this design, C_p is around 6.4 pF and C_{pp} is chosen to be 0.8 pF, which occupies 1000 μm^2 with poly-poly implementation. The switches are implemented with transmission gates; transistor dimensions are 60/0.4 and 20/0.4 for PMOS and NMOS, respectively. The current flowing through the OFF switchable current source I_{OFF} is around 240 μA and ΔV_{gate} is around 200 mV. Notice that the data D drives an equivalent

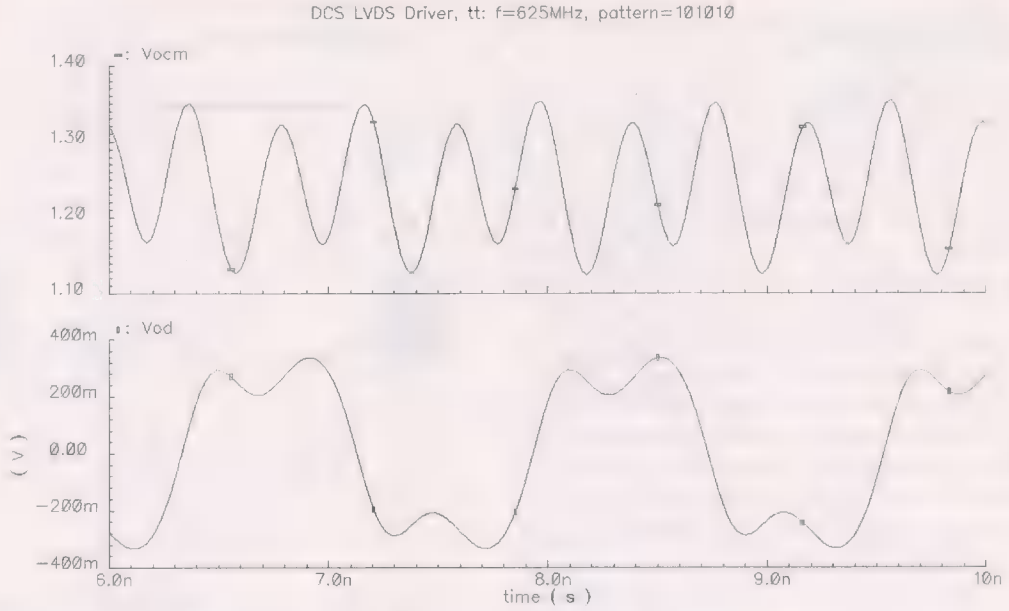


Fig. 10. Common-mode and differential-mode DCS LVDS driver output waveforms with load model.

capacitance of approximately 0.7 pF; hence D and $\overline{D}$ are not severely affected by the pull up/down capacitor C_{pp}.

When the switchable current source M1 (M2) is turned on, the pull up/down capacitor C_{pp} is connected to ground (logic ZERO); so it is important to reduce the substrate noise to minimize its effect on the output signal amplitude. When M1 (M2) is turned off, C_{pp} is connected to the power supply (logic ONE). Since M1 (M2) is working in the subthreshold region, its current is very small hence the supply variation has very limited effect on the output signal amplitude.

Compared to the active pull up/down circuit, this passive pull up/down circuit is faster as a result of the capacitors used, consumes less power, and the up/down voltage changes are symmetrical. With symmetrical voltage changes, the switches S1 and S2 can be small and the speed of the Buf-A is relaxed. Also, the driver's architecture is simpler and, therefore, more robust.

D. Simulation Results

The transistor dimensions of the DCS and SCS LVDS driver cores are shown in Table I. The simulated DCS LVDS driver output common-mode and differential-mode voltages with data rate of 1.25 Gb/s are shown in Fig. 10. In this simulation, the models of the electrical static discharge (ESD) device, bonding wire, and package are included. Also, the termination resistor and load capacitors at the receiver end are included. Notice that both common-mode and differential-mode output voltages are within the LVDS standard specifications.

From the discussions in the aforementioned sections, it can be seen that the key design issue of the SCS LVDS driver is to control the switchable current source gate voltage $V_{\rm gate}$ and so the corresponding drain current. Fig. 11 shows the simulation results for the switchable current source gate voltage $V_{\rm gate}$ (top trace), transistor drain current I_D (middle trace) and the corresponding output differential voltage (bottom trace); the load model was simplified in order to see $V_{\rm gate}$ change more clearly. Notice that the gate voltage $V_{\rm gate}$ and the corresponding drain current I_D switches properly. The transition time is only around 240 ps and it can be seen that the rising time and falling time of the output signal are within the specifications (300–500 ps). The small transition time is mainly due to the passive capacitors used for the pull up/down circuit, and operating the switchable current sources in a subthreshold region when they are turned OFF. The gate voltage variation $\Delta V_{\rm gate}$ is around 200 mV, and the drain current $I_{\rm ON}$ and $I_{\rm OFF}$ are around 6.4 mA and 240 μA, respectively. Notice that the gate voltage $V_{\rm gate}$ and the drain current I_D present small variations. They are due to the transients of charging/discharging the parasitic capacitances.

TABLE I
TRANSISTOR DIMENSIONS OF THE DCS AND SCS LVDS CORES

Transistor	M1=M2	M3=M4	M5
DCS LVDS W/L (μm/μm)	4000/.4	600/.4	2000/.4
SCS LVDS W/L (μm/μm)	4000/.4	200/.4	1000/.4

IV. Experimental Results

Both the DCS and SCS LVDS drivers have been fabricated in the TSMC 0.35-μm CMOS process through the MOSIS service; the active die areas are 0.11 mm^2 and 0.14 mm^2, respectively. The chip micrograph is shown in Fig. 12 and was packaged in a 64-pin ceramic quad flat package. According to the experimental results, the DCS LVDS driver operates properly for a data rate up to 1.4 Gb/s and the SCS LVDS driver operates for data rates up to 1.2 Gb/s. Those shortcomings might be alleviated if more advanced processes or N-type switchable current sources are used.

Figs. 13 and 14 show the DCS LVDS driver differential output eye diagrams with $2^{31}-1$ pseudorandom bit sequence (PRBS) pattern and data rates of 680 Mb/s and 1.0 Gb/s, respectively. The single-ended output signal swings are around 340 mV and

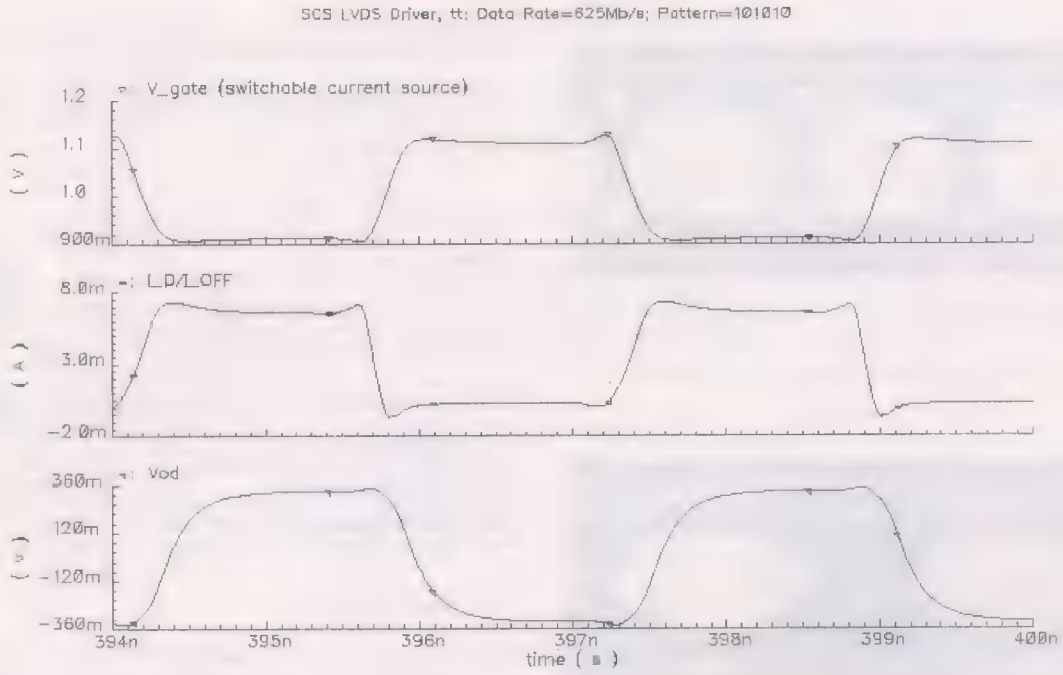


Fig. 11. Switchable current source gate voltage (top), drain current (middle), and the output differential voltage (bottom).

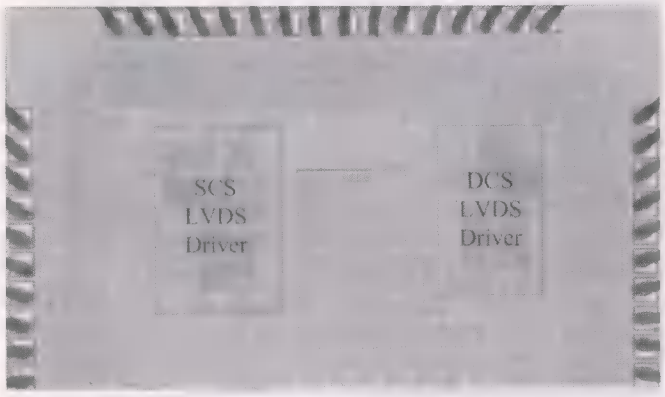


Fig. 12. DCS and SCS LVDS drivers chip micrograph.

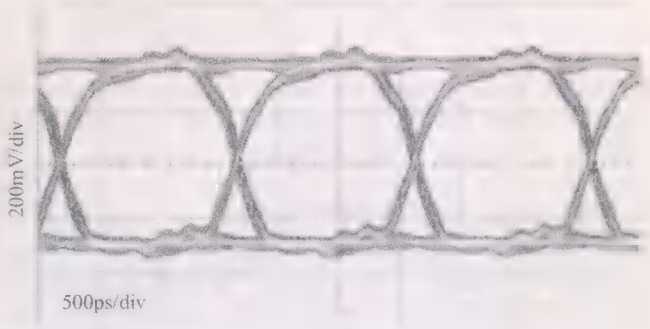


Fig. 13. DCS LVDS driver eye diagram (data rate = 680 Mb/s).

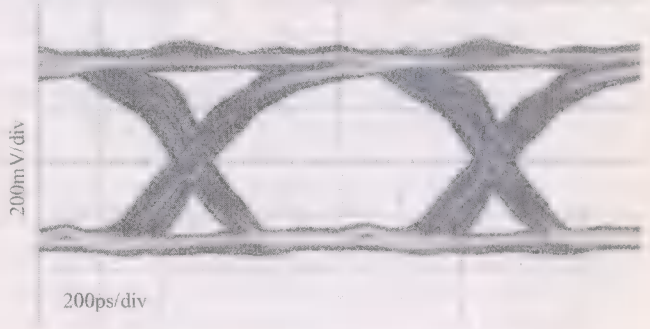


Fig. 14. DCS LVDS driver eye diagram (data rate = 1.0 Gb/s).

the measured root-mean-square (RMS) jitters are 15 and 36 ps, respectively. The eye openings are 90% and 80%, respectively. Figs. 15 and 16 show the SCS LVDS driver differential eye diagram with $2^{31} - 1$ PRBS at data rates of 680 Mb/s and 1.0 Gb/s, respectively. The differential output signal swings are 680 mV and the measured RMS jitters are 28 and 50 ps, respectively. The eye openings are 85% and 60%, respectively.

Compared to the DCS LVDS driver, the SCS LVDS driver presents larger jitter and narrower open eyes. Several factors contribute to this. First, the rising and falling times of the SCS LVDS driver output signal are larger than those of the DCS LVDS driver output signal, which is due to the finite transition times of the gate voltage and drain current of the switchable current sources. Second, while the drain current of the PMOS current sources in the DCS LVDS driver remains constant, the drain current of the switchable current sources presents some variations, which is due to the transients of charging/discharging the parasitic capacitances. Also, the effect of the charge injection on the driver's output nodes is more pronounced for the SCS LVDS driver than for the DCS LVDS driver.

The total current consumption (including both static and dynamic) of the two LVDS structures for different data rates are given in Table II. The dynamic power consumed by the parasitic capacitance of the NMOS switches has been neglected for both structures. While in this table the current consumption of the DCS LVDS driver only consists the static tail current, that of the

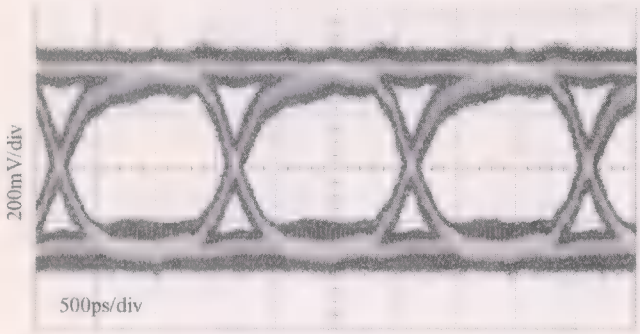


Fig. 15. SCS LVDS driver eye diagram (data rate = 680 Mb/s).

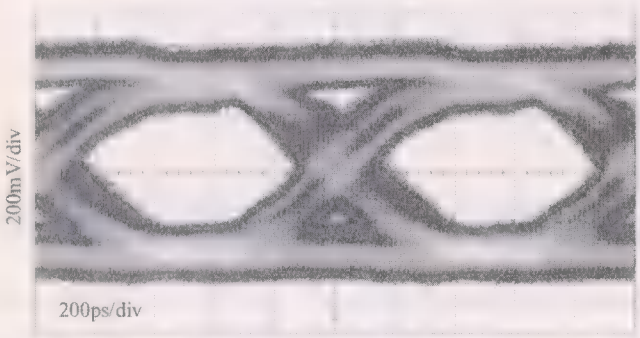


Fig. 16. SCS LVDS driver eye diagram (data rate = 1.0 Gb/s).

SCS LVDS driver includes the current drawn by the buffer-connected amplifier Buf-A, the dynamic current consumed by the parasitic capacitance of the switchable current sources, and the static tail current. It can be seen that the SCS LVDS driver draws much less current than the DCS LVDS driver.

A comparison among these two structures and a previously reported LVDS driver [8] is shown in Table III. This reported driver is based on typical LVDS configurations, except that it uses all NMOS switches to reduce the charge injection effects. Another reported LVDS driver requires an external resistor and two reference voltages [9]. Notice that both the DCS and SCS LVDS drivers consume less power than previous realizations. Especially for the SCS LVDS driver, by dynamically switching the current sources, it reduces the power consumption by 60% compared to the previous implementations (if the same signal swing is maintained). In addition, while the previously reported LVDS drivers cannot operate properly with low-voltage supplies, both the DCS and SCS LVDS drivers are suitable for low-voltage supply applications, and they are still compliant to LVDS standards and operate properly at very high data rates.

In addition to the low-power consumption, the other benefits of the low-voltage supply drivers are reduced EMI and costs related to the packaging and cooling systems. Being able to operate with low-voltage supplies makes it possible to use the same supply for both the core circuits and the I/O drivers, which can simplify both circuit and PCB design.

V. Conclusion

Two LVDS driver structures suitable for very low-voltage supplies (as low as 1.8 V) are discussed. The DCS LVDS driver is simple and fast. Despite the dynamic power consumed by the parasitic capacitance of NMOS switches, the DCS LVDS driver power consumption is almost constant, regardless of the data patterns. A drawback of the DCS LVDS driver is that its static current consumption is twice the minimum required by the output voltage swing. Another drawback is that the transistor dimension of the switches and the bottom NMOS current sources are relatively large because of the larger amount of current used, therefore die area and parasitic capacitors increase.

The SCS LVDS driver is more complex compared to the DCS LVDS driver, but its most significant advantage is that the static current consumption is kept to the minimum as required by the output voltage swing and load. Since it is needed to charge/discharge the parasitic capacitance associated with the switchable current sources, the SCS LVDS driver power consumption depends on the data pattern, even if we neglect the dynamic power consumed by the parasitic capacitance of NMOS switches. The higher the data rate, the larger the dynamic power consumption of the pull up/down circuit is.

TABLE II
Current Consumption for DCS and SCS LVDS Drivers

Data Rate (Mb/s)	680	1000
DCS $I_{average}$ (mA)	12.8	12.8
SCS $I_{average}$ (mA)	8.5	9.0

TABLE III
Comparison With Previous Realizations

	[8]	DCS	SCS
Technology	0.35μm CMOS	0.35μm CMOS	0.35μm CMOS
Output Voltage Swing (mV)	412	340	340
Maximum Data Rate (Mb/s)	1200	1400	1200
Static Power Consumption (mW)	43	23	12.8
Cell Size (mm^2)	0.17	0.11	0.14
Supply Voltage (V)	3.3	1.8	1.8

References

[1] R. A. Nordin, A. F. J. Levi, R. N. Nottenburg, J. O'Gorman, T. Tanbun-Ek, and R. A. Logan, "A systems perspective on digital interconnection technology," *J. Lightwave Technol.*, vol. 10, pp. 811–827, Jun. 1992.

[2] *F100K ECL 300 Series Data Book and Design Guide*. National Semiconductor Corp., 1992.

[3] *IEEE Standard for Low-Voltage Differential Signals (LVDS) for Scalable Coherent Interface (SCI), 1596.3 SCI-LVDS Standard*, IEEE Std 1596.3-1996, 1996.

[4] *IEEE Standard for Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications-Media Access Control (MAC) Parameters, Physical Layer and Management Parameters for 10 Gb/s Operation*, IEEE Draft P802.3ae/D5.0, May 1, 2002.

[5] H. W. Johnson and M. Graham, *High-Speed Digital Design, A Handbook of Black Magic*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[6] R. Senthinathan and J. L. Prince, "Application specific CMOS output driver circuit design techniques to reduce simultaneous switching noise," *IEEE J. Solid-State Circuits*, vol. 28, no. 12, pp. 1383–1388, Dec. 1993.

[7] Y. Ohtomo, M. Nogawa, and M. Ino, "A 2.6-Gbps/pin SIMOX-CMOS low-voltage-swing interface circuit," *IEICE Trans. Electron.*, vol. E79-C, no. 4, pp. 524–529, Apr. 1996.

[8] A. Boni, A. Pierazzi, and D. Vecchi, "LVDS I/O interface for Gb/s-per-Pin operation in 0.35-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 706–711, Apr. 2001.

[9] T. Gabara, W. Fisher, W. Werner, S. Siegel, M. Kothandaraman, P. Metz, and D. Gradl, "LVDS I/O buffers with a controlled reference circuit," in *Proc. ASIC Conf.*, Sep. 1997, pp. 311–315.

Mingdeng Chen (S'01–M'04) was born in Jingzhou, Hubei, China. He received the B.S. degree in applied mathematics and M.S. degree in aerospace engineering, both from National University of Defense Technology, in 1993 and 1996, respectively, and the Ph.D. degree from Texas A&M University, College Station, in 2003.

He has been with Agere Systems, Allentown, PA, as an IC Design Engineer, since 2003. He has been involved in mixed-signal circuit design for hard disk driver read channels. He worked on continuous-time filter design and high-speed serial interface design, as an intern IC Designer, at RocketChips, and Communication Technology Division, Xilinx, Austin, TX, in 2000 and 2002, respectively. His research interests include analog/RF, and mixed-signal circuit design.

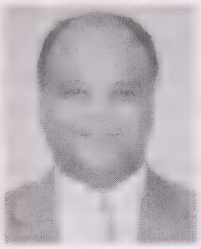
Jose Silva-Martinez (SM'98) was born in Tecamachalco, Puebla, México. He received the B.S. degree in electronics from the Universidad Autónoma de Puebla, in 1979, the M.Sc. degree from the Instituto Nacional de Astrofísica Optica y Electrónica (INAOE), Puebla, in 1981, and the Ph.D. degree from the Katholieke Univesiteit Leuven, Leuven, Belgium, in 1992.

From 1981 to 1983, he was with the Electrical Engineering Department, INAOE, where he was involved with switched-capacitor circuit design. In 1983, he joined the Department of Electrical Engineering, Universidad Autónoma de Puebla, where he remained until 1993; He was a co-founder of the graduate program on Opto-Electronics in 1992. From 1985 to 1986, he was a Visiting Scholar in the Electrical Engineering Department, Texas A&M University. In 1993, he rejoined the Electronics Department, INAOE, and from May 1995 to December 1998, he was the Head of the Electronics Department; he was a co-founder of the Ph.D. program on Electronics in 1993. He is currently with the Department of Electrical Engineering (Analog and Mixed Signal Center) Texas A&M University, College Station, where he is an Associate Professor. His current field of research is in the design and fabrication of integrated circuits for communication and biomedical applications.

Dr. Silva-Martínez has served as IEEE CASS Vice President Region 9 (1997–1998), and as Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART II during 1997–1998 and May 2002–December 2003. Since January 2004 he is serving as Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART I. He was the main organizer of the 1998 and 1999 International IEEE CAS Tour in region 9, and Chairman of the International Workshop on Mixed-Mode IC Design and Applications (1997–1999). He is the inaugural holder of the TI Professorship-I in Analog Engineering, Texas A&M University. He was a co-recipient of the 1990 European Solid-State Circuits Conference Best Paper Award

Michael Nix received the B.S.E.E. degree from Texas A&M University in 1976.

From 1976 to 1978, he was in Fortran programming for Lockheed Electronics, working on the Space Vehicle Dynamics Simulator. From 1978 to 1979, he was doing board-level design for Sperry Avionics, and worked on auto-pilots for business jets. From 1979 to 1983, he was with the Integrated Circuit Design Group of Mostek, and from 1983 to 1987, with integrated circuit design for Texas Micro Engineering/Crystal Semiconductor, where he was dealing with analog, digital, and mixed-signal design for a variety of products in CMOS processes. From 1987 to 2000, he was doing integrated circuit design for Advanced Micro Devices. Some of the projects he has dealt with include voice CODECs in CMOS processes from 1 to 0.35 microns. Since 2000, he has been with RocketChips/Xilinx, dealing with mixed-signal design for data conversion devices and SERDES in CMOS processes from 0.35 to 0.09 microns. He has 15 U.S. patents granted, and three pending.



Moises E. Robinson (S'87–M'91) received the B.S. (*summa cum laude*) and M.S. degrees in electrical engineering from Texas A&M University in 1989 and 1991, respectively.

From 1991 to 1994, he was an Analog/Mixed-Signal IC Designer with IMP, Pleasanton, CA, where he was involved in the design of high-speed circuits for disk-drive applications. From 1994 to 1996, he was a Senior Design Engineer for Crystal Semiconductors, where he was involved in the development of delta-sigma data converters. From 1996 to 1998, he was a Senior Analog/Mixed-Signal Designer for Oak Technology, working on Audio and Modem CODEC products for the AC97 Audio Standard. Since 1998, he has been a Technical Director with the Communications Technology Division of Xilinx, Austin, TX (formerly RocketChips). He has published more than 20 journal and conference papers, and has more than ten issued U.S. patents in the area of mixed-signal circuit design. His current research interests include high-speed serial communications and low-noise clock generation.

High-Performance Low-Power Dual Transition Preferentially Sized (DTPS) Logic

Woopyo Jeong and Kaushik Roy, *Fellow, IEEE*

***Abstract*—We present a dual transition preferentially sized (DTPS) logic that uses two separate paths—one for the fast propagation of low-to-high signal and the other for fast propagation of high-to-low signal. DTPS logic is suitable for multistage buffers and critical sections of datapaths requiring good noise immunity and low power dissipation while achieving high performance. We derived formulas to obtain optimal tapering factors of multistage buffers based on preferentially sized (PS) inverters, and implemented DTPS logic using the optimal tapering factors. We fabricated datapaths based on static CMOS logic, domino logic, and DTPS logic in 0.18-μm technology. DTPS logic shows 15% and 16% improvements in performance and power dissipation, respectively, over domino, and 42% improvement in performance compared to static CMOS.**

***Index Terms*—Dual transition preferentially sized (DTPS), preferentially sized logic, tapering factor.**

I. Introduction

With the scaling of process technology, high performance and low power consumption are becoming important issues in circuit design. The use of domino circuits is one way to alleviate the problem of high-performance circuit design. However, domino circuits consume more power than standard CMOS logic, and are susceptible to noise (for scaled technologies with low transistor threshold voltage) because in the evaluation mode intermediate nodes may be floating [1]–[3].

In order to achieve good noise immunity and low power consumption while achieving performance comparable to domino logic, we propose dual transition preferentially sized (DTPS) logic, which consists of dual monotonic datapaths (one is fast for rising transition of input, and the other is for falling transition) using preferentially sized (PS) circuits [4]. Since a PS inverter chain uses up-sized inverters and down-sized inverters alternately to speed up data propagation in evaluation cycle, the ratio of output capacitance to input capacitance of even stages of multistage PS buffers is different from that of odd stages. Hence, different tapering factors should be used for even and odd stages, which are also different from the tapering factor of normal inverter chains. We derive formulas for optimal tapering factors of multistage buffers based on PS inverters to minimize the propagation delay. DTPS is implemented based on PS inverter chains using dual tapering factors. DTPS logic is not only suitable for multistage buffers but also ideal for critical sections of datapaths requiring high performance and low power consumption. We also describe how to design DTPS logic using a high sizing ratio in critical paths of design to achieve a very high performance. We fabricated datapaths based on static CMOS logic, domino logic, and DTPS logic. The measurement results show the advantages of DTPS logic.

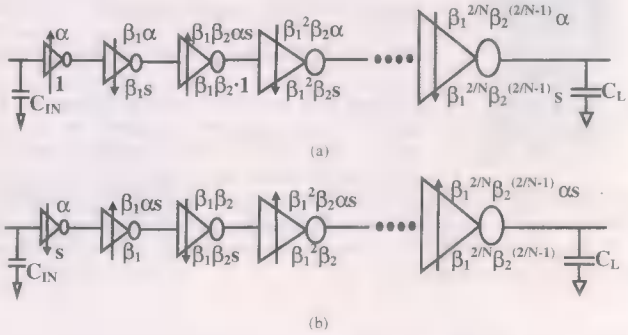


Fig. 1. N-stage preferentially sized buffers with dual tapering factors (a) starting an up-sized inverter and (b) starting a down-sized inverter.

II. Preferentially Sized (PS) Logic

In order to design high performance multistage buffers and datapaths using DTPS logic proposed in this paper, we first consider PS buffers that are the building blocks of the DTPS buffers. Then, in order to minimize the delay due to the PS buffers, optimal tapering factors are considered, which are different from the tapering factor of normal inverter chains [5], [6]. Fig. 1 shows some examples of how to adjust the sizes of PS circuit style, where s is the sizing ratio and α is the ratio of optimal size of the PMOS to NMOS in a static CMOS inverter. β is the tapering factor, which is the ratio of output capacitance to the input capacitance of an inverter. The arrows represent the sizing directions of the inverters. In this paper we used sizing ratios greater than 1. Since a multistage PS buffer uses up-sized inverters and down-sized inverters alternately, we should use two tapering factors for a multistage PS buffer—one for the even stages and the other for the odd stages. In Fig. 1, the N-stage PS buffer uses two tapering factors (β_1 and β_2). One tapering factor (β_1) is used for the even stages, and the other (β_2) is for the odd stages. Hence, the output capacitive load, C_L, is $C_L = (\beta_1\beta_2)^{(N/2)}C_{\mathrm{IN}} = (\beta_1\beta_2)^{(N/2)}(1+\alpha\cdot s)C_{g0}$, where C_{go} is the input gate capacitance per unit area.

Manuscript received February 5, 2004; revised August 1, 2004.

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: jeongw@ecn.purdue.edu; kaushik@ecn.purdue.edu).

Digital Object Identifier 10.1109/JSSC.2004.841040

In Fig. 1(a), the low-to-high propagation delay of the first stage (t_{PLH1}) and the high-to-low propagation delay of the second stage (t_{PHL2}) are given as follows:

$$t_{\mathrm{PLH1}} = \frac{V_{\mathrm{DD}}}{(V_{\mathrm{DD}} - V_{tp})^2}[(1+\alpha\cdot s)C_{d0} + \beta_1(\alpha+s)C_{g0}]\frac{t_{\mathrm{ox}}}{\mu_p\cdot\varepsilon_{\mathrm{ox}}}\cdot\frac{L}{\alpha\cdot s} \quad (1)$$

$$t_{\mathrm{PHL2}} = \frac{V_{\mathrm{DD}}}{(V_{\mathrm{DD}} - V_{tn})^2}[(\alpha+s)C_{d0}\cdot\beta_1 + (1+\alpha\cdot s)C_{g0}\cdot\beta_1\cdot\beta_2]\frac{t_{\mathrm{ox}}}{\mu_n\cdot\varepsilon_{\mathrm{ox}}}\cdot\frac{L}{\beta_1\cdot s}. \quad (2)$$

The total propagation delay is minimum when the propagation delays of each stage are same ($t_{\mathrm{PHL1}} = t_{\mathrm{PLH2}} = t_{\mathrm{PHL3}}\ldots$) [2]. Hence, we can obtain $\beta_2 \approx (\alpha+s)/(1+\alpha\cdot s)\cdot\beta_1$ from (1) and (2), and the total propagation delay ($t_{\mathrm{PHL1}} + t_{\mathrm{PLH2}} + t_{\mathrm{PHL3}} + \cdots + t_{\mathrm{PHL}N}$) can be written as follows:

$$t_p = N\cdot\left(\frac{C_L}{C_{\mathrm{IN}}}\right)^{\frac{1}{N}}\cdot\frac{V_{\mathrm{DD}}}{(V_{\mathrm{DD}}-V_t)^2}\cdot\sqrt{(1+\alpha\cdot s)(\alpha+s)}\cdot C_{g0}\cdot\frac{t_{\mathrm{ox}}}{\mu_n\cdot\varepsilon_{\mathrm{ox}}}\cdot\frac{L}{s}. \quad (3)$$

In (3), when $s = 1$, t_{p} is the total propagation delay of an N-stage normal multistage buffer. An optimal number of stages of multistage PS buffers with dual tapering factors (N_{opt}), which is obtained by solving $\partial t_{\mathrm{p}}/\partial N = 0$, is $\ln(C_L/C_{\mathrm{IN}})$. Hence, we can get $\beta_1\beta_2 = (C_L/C_{\mathrm{IN}})^{(2/N)} = (e^N)^{(2/N)} = e^2$ from $C_L = (\beta_1\beta_2)^{(N/2)}C_{\mathrm{IN}}$. Since $\beta_2 \approx (\alpha+s)/(1+\alpha\cdot s)\cdot\beta_1$

$$\beta_1 = \sqrt{\frac{1+\alpha\cdot s}{\alpha+s}}e \qquad \beta_2 = \sqrt{\frac{\alpha+s}{1+\alpha\cdot s}}e \quad (4)$$

which are different from the optimal tapering factors of normal multistage buffers.

Fig. 2 depicts delays of multistage PS buffers with dual tapering factors, which are normalized to the delay of a normal multistage buffer (when $s = 1$). The dotted lines represent simulation results for different number of stages, and the solid line shows the analytical result obtained from (3). In Fig. 2, when the sizing ratio is 5, the delays of multistage PS buffers are 56% less than those of normal multistage buffers. However, since the propagation delay in the precharge cycle is much larger than that of the evaluation cycle, conventional PS logic does require a clock signal, though only selective logic gates may require the clock to reset the PS logic in the precharge cycle [4]. This increases the clock load and, hence, the power consumption.

III. Dual Transition Preferentially Sized (DTPS) Logic

We propose to use DTPS logic, in which the sizes of PS inverters on each datapath are determined based on the optimal tapering factors, to achieve high performance and low power dissipation. DTPS logic does not require a clock signal. Fig. 3 shows an example of DTPS logic that achieves high performance by duplicating signal paths: both paths consist of PS logic, in which one signal path is for fast rising transition of

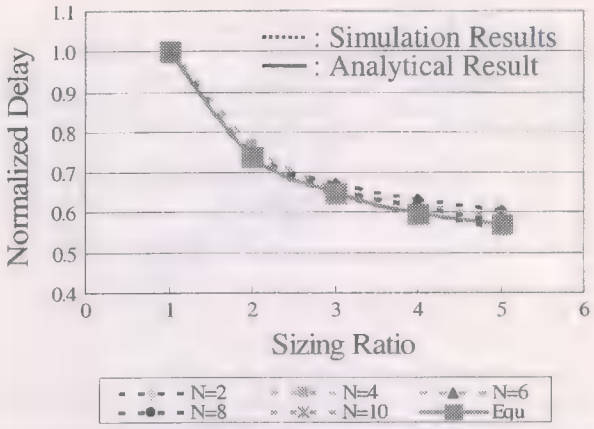


Fig. 2. Normalized delay of preferentially sized buffer using dual tapering factors.

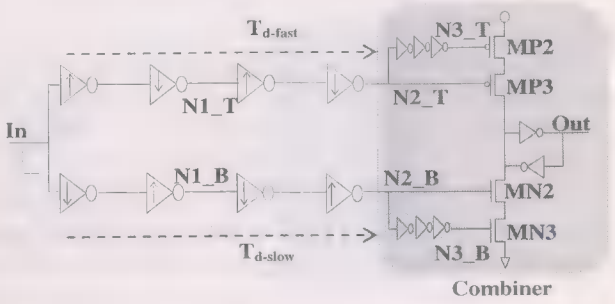


Fig. 3. Four-stage DTPS buffer with low sizing ratio.

input, while the other is for fast falling transition. A combiner detects the earliest transition, latches it, and then transfers the data to the next stage. Hence, DTPS logic can achieve fast propagation delay both in evaluation and precharge modes. For example, in Fig. 3, if the input toggles from high-to-low, the top path is faster than the bottom path. Hence, though both nodes N2_T and N2_B transit from high to low, N2_T transits faster than the node N2_B. The high-to-low transition on N2_T turns on MP3, while N3_T stays at low, which makes output transit from high to low. If the input toggles from low to high, the node N2_B transits from low to high faster than the node N2_T. The low-to-high transition on N2_B turns on MN2, while N3_B stays at high, which makes output transit from low to high.

The circuit diagram shown in Fig. 3 is valid only when low sizing ratio (s) is used, in which the difference between fast propagation delay and slow propagation delay is less than the clock period. To achieve high performance, highly preferentially sized inverters may be required. However, using highly preferentially sized inverters (for a certain sizing ratio for which $T_{\mathrm{p\_slow}} \geq T_{\mathrm{p\_fast}} + T_{\mathrm{cycle}}$) can make the transition of slow data due to previous input signal and the transition of fast data due to the current input signal occur at almost the same time at a certain node, creating a glitch (spurious transition).

Fig. 4 shows an example of this functional problem of DTPS buffers with a high sizing ratio. The previous input, IN[i-1], toggles from low to high, and the current input, IN[i], toggles from high to low. Propagation due to current input (IN[i]) is faster than due to previous input (IN[i-1]) on the top datapath in Fig. 3 because transition directions of PS inverters due to

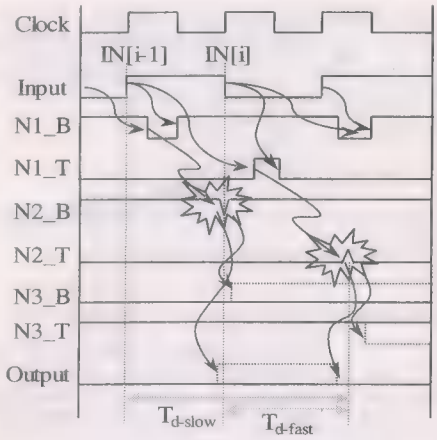


Fig. 4. Timing diagram of the DTPS buffer having high sizing ratio ($T_{\text{d\_slow}} \geq T_{\text{d\_fast}} + T_{\text{cycle}}$).

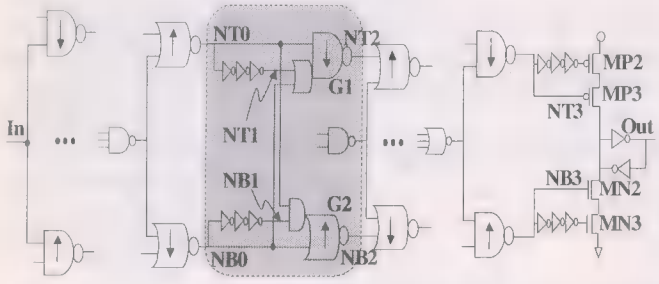


Fig. 5. Cross-path DTPS logic applied for critical sections of datapaths.

the current input are the same as their sizing directions. Propagation due to previous input (IN[i-1]) is slow on the top path, and hence, fast high-to-low transition due to current input and slow low-to-high transition due to the previous input occurs simultaneously at N2_T. It produces a glitch at N2_T and can not turn on MP3. Hence, even though the input toggles from high-to-low, the output does not toggle while keeping the previous data (high). This problem occurs when the delay of the slow path is larger than summation the delay of fast path and cycle time ($T_{\text{p\_slow}} > T_{\text{p\_fast}} + T_{\text{cycle}}$).

To solve the spurious transition problem of the multistage DTPS buffer having highly preferentially sized inverters, we propose a multistage cross-path DTPS buffer that uses extra logic to take care of the robustness problem by reducing the propagation delay in the slow path. The proposed cross-path DTPS circuit techniques are applicable to multistage buffers and critical sections of datapaths requiring very high performance with low power consumption. Fig. 5 shows the proposed cross-path DTPS logic applied to critical sections of the datapaths requiring very high performance. The compound gates (G1 and G2) are added to handle the robustness problem of DTPS logic mentioned above. The architecture of DTPS in a datapath is the same as that in the multistage DTPS buffer except that a datapath consists of combinational gates like NAND, NOR, or other complex gates. We can partition combinational gates on each datapath into two parts: gates having a critical input signal and gates having noncritical input signals. For example, in Fig. 6, a 4-input NAND gate (G1) on the top datapath can be partitioned into 2-input NAND gate (G3) and 3-input NOR gate

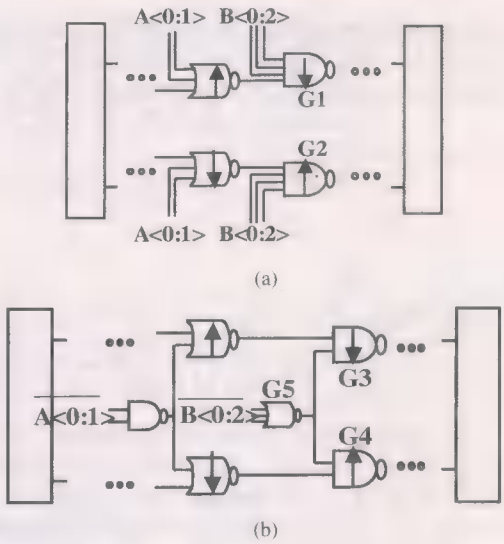


Fig. 6. DTPS logic for datapaths (a) before logic restructuring and (b) after logic restructuring.

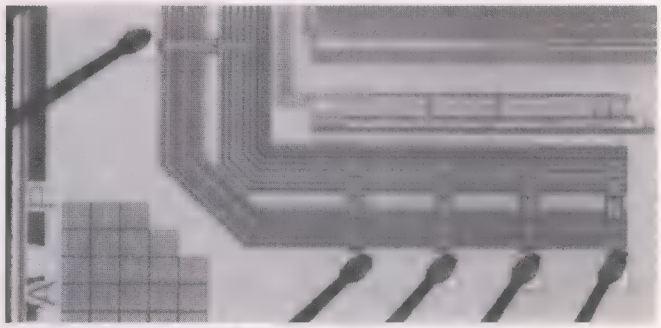


Fig. 7. Chip microphotograph.

(G5), and G2 on bottom datapath can also be partitioned into G4 and G5. Since G5 is common, it can be shared as shown in Fig. 6(b). Gates G3 and G4 are on the critical paths, however, G5 is on the noncritical path. Logic restructuring reduces gate fan-in and the size of transistors on the critical paths, which decreases load capacitance on the critical paths, thereby reducing the delay and the layout area [2].

In Fig. 5, when the input transits from low to high, data propagation through the top path is faster than through the bottom one. Hence, NT0 goes high faster than NB0 while NT1 stays at high, which makes NT2 transit from high to low. Low-to-high transition of NT0 also makes NB2 transit from high to low independent of the data at node NB0, i.e., high-to-low transition of NB2 occurs before low-to-high transition of NB0. In this case, NT1 should transit from high to low after NB0 transits from low to high. If NT1 transits before NB0 transition, a glitch occurs at NT2. On the other hand, when input transits from high to low, NT2 on the top path is determined by the fast high-to-low transition of NB0, while NT1 is low. The delay of the slow path of this DTPS buffer, $T'_{\text{d\_slow}}$, is defined as $0.5 \cdot (T_{\text{d\_fast}} + T_{\text{d\_slow}}) + T_{\text{comp}}$, where T_{comp} is the delay due to a compound gate ($T'_{\text{d\_slow}} < T_{\text{d\_slow}}$). Hence, inserting extra component gates for determining slow path can reduce the delay of the slow path and remove the glitch problem of DTPS logic with high sizing ratio mentioned earlier.

TABLE I
MEASUREMENT RESULTS (AT 100 MHZ)

Process / Supply Voltage	0.18 μm CMOS		
Supply Voltage	1.8V		
	DTPS	Domino	Static CMOS
Delay	5.5ns	6.5ns	9.5ns
Power Consumption	0.238mW	0.284mW	0.126mW
Layout Area	11466um^2	7357um^2	6200um^2

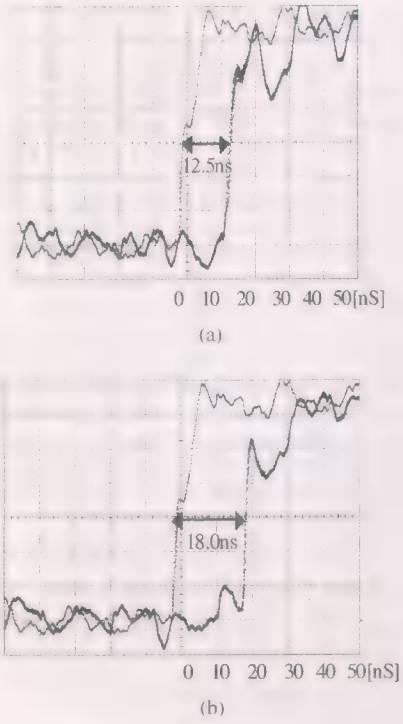


Fig. 8. Measured output waveforms of (a) bypass and (b) DTPS logic.

IV. EXPERIMENTAL RESULTS

We fabricated datapaths based on DTPS, domino, and static CMOS logic using TSMC 0.18-μm CMOS technology, and compared DTPS logic with domino logic and static CMOS logic with respect to performance and power consumption. Fig. 7 shows the chip microphotograph. It consists of a bypassing path, three datapaths based on DTPS logic, domino logic, and static CMOS, and multiplexeers (muxes) to select one of the datapaths and bypass path. Fig. 8 shows that the measured delays of the datapath based on DTPS logic and the bypass path are 18.0 and 12.5 ns, respectively. Hence, the real delay of DTPS logic is 5.5 ns. Using the same method, the delays of datapaths based on domino logic and static CMOS logic are obtained.

Table I summarizes the measured delays and power consumptions of datapaths of different logic styles. It shows 15% and 16% improvements in performance and power, respectively, over domino logic. DTPS and domino logic show 42% and 31% delay improvements over the static CMOS logic.

V. CONCLUSION

In this paper we proposed DTPS logic, which is suitable for multistage buffers and critical sections of datapaths requiring a very high performance with low power consumption. We derived expressions for optimal tapering factors of multistage buffers based on PS inverters to minimize the propagation delay. Analytical results show that PS buffers with dual tapering factors can achieve up to 13% performance improvement over ones using one tapering factor. For the PS buffers using dual tapering factors, the difference between the analytical results and the simulation results is less than 10%. We fabricated test chip for datapaths based on DTPS logic, static CMOS, and Domino logic. The measured results show 15% and 16% improvements in performance and power, respectively, over Domino and 42% delay improvement over the static CMOS logic.

REFERENCES

[1] R. Krambeck, C. M. Lee, and H.-F. S. Law, "High-speed compact circuits with CMOS," *IEEE J. Solid-State Circuits*, vol. SC-17, no. 3, pp. 614–619, Jun. 1982.

[2] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*, New Jersey: Prentice Hall, 1996.

[3] P. Larsson and C. Svensson, "Noise in digital dynamic CMOS circuits," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 655–662, Jun. 1994.

[4] A. Solomatmikov, D. Somasekhar, and K. Roy, "Skewed CMOS: Noise-tolerant high-performance low-power static circuits family," *IEEE Trans. VLSI Syst.*, vol. 10, no. 4, pp. 469–476, Aug. 2002.

[5] N. Hedenstierna and K. O. Jeppson, "CMOS circuit speed and buffer optimization," *IEEE Trans. Computer-Aided Des.*, vol. CAD-6, no. 2, pp. 270–281, Mar. 1987.

[6] C. Prunty and L. Gal, "Optimum tapered buffer," *IEEE J. Solid-State Circuits*, vol. 27, no. 1, pp. 118–119, Jan. 1992.

Woopyo Jeong received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 1991 and 1993, respectively. He is currently working toward the Ph.D. degree in electrical and computer engineering at Purdue University, West Lafayette, IN.

In 1993, he joined Samsung Electronics Company, Ltd., Korea, where he was engaged in research and development for EDO, synchronous DRAM, and Rambus DRAM. His research interests include high-performance and low-power circuit design.

Kaushik Roy (SM'95–F'01) received the B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1990.

He was with the Semiconductor Process and Design Center, Texas Instruments Inc., Dallas, TX, where he worked on FPGA architecture development and low-power circuit design. He joined the electrical and computer engineering faculty at Purdue University, West Lafayette, IN, in 1993, where he is currently a Professor. He is on the Technical Advisory Board of Zenasis Inc. and a Research Visionary Board Member of Motorola Laboratories. He has published more than 250 papers in refereed journals and conferences, holds six patents, and is ■ coauthor of a book on low-power CMOS VLSI design. His research interests include VLSI design/CAD with particular emphasis in low-power electronics for portable computing and wireless communications, VLSI testing and verification, and reconfigurable computing.

Dr. Roy received the National Science Foundation Career Development Award in 1995, the IBM Faculty Partnership Award, the AT&T/Lucent Foundation Award, Best Paper Awards at the 1997 International Test Conference, IEEE 2000 International Symposium on Quality of IC Design, and IEEE Latin American Test Workshop, and is currently a Purdue University Faculty Scholar Professor. He has been on the editorial board of IEEE *Design and Test*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and IEEE TRANSACTIONS ON VLSI SYSTEMS. He was Guest Editor for the Special Issue on Low-Power VLSI in the IEEE *Design and Test* (1994), IEEE TRANSACTIONS ON VLSI SYSTEMS (June 2000), and IEE *Proceedings—Computers and Digital Techniques* (July 2002).

Design Considerations for Soft Embedded Programmable Logic Cores

Steven J. E. Wilton, *Senior Member, IEEE*, Noha Kafafi, James C. H. Wu, *Member, IEEE*, Kimberly A. Bozman, Victor O. Aken'Ova, and Resve Saleh, *Senior Member, IEEE*

***Abstract*—As integrated circuits become increasingly more complex and expensive, the ability to make post-fabrication changes will become much more attractive. This ability can be realized using programmable logic cores. Currently, such cores are available from vendors in the form of "hard" rectangular layouts. In this paper, we focus on an alternative approach for fine-grain programmability: vendors supply a synthesizable RTL version of their programmable logic core (a "soft" core) and the integrated circuit designer synthesizes the programmable logic fabric using standard cells. Although this technique suffers in terms of speed, density, and power overhead, the task of integrating such cores is far easier than the task of integrating "hard" cores into an ASIC or SoC. When the required amount of programmable logic is small, this ease of use may be more important than the increased overhead. This paper presents two synthesizable "soft" programmable logic core architectures and describes their associated place and route issues. We compare the two architectures to each other, and to a "hard" programmable logic core. We also show how these cores can be made more efficient by creating a nonrectangular architecture, an option not usually available to "hard" core vendors. Finally, a proof-of-concept integrated circuit containing one of these cores is described.**

***Index Terms*—Field-programmable gate arrays, programmable logic, SoC design.**

I. Introduction

RECENTLY, we have witnessed impressive improvements in the achievable density of integrated circuits. In order to maintain this rate of improvement, designers need new techniques to manage the increased complexity inherent in these large chips. One such emerging technique is the system-on-a-chip (SoC) design methodology. In this methodology, pre-designed and pre-verified blocks, often called cores or intellectual property (IP), are obtained from internal sources or third-parties, and combined on a single chip. These cores may include embedded processors, memory blocks, interface blocks and components that handle application specific processing functions. Large productivity gains can be achieved using this approach. In fact, rather than implementing each of these components separately, the role of the SoC designer is to integrate them onto a chip to implement complex functions in a relatively short amount of time.

One major issue today in SoC design is the overall design cost in terms of engineering costs, the cost of IP blocks and the rising costs of masks in advanced technologies. For this reason, it is desirable to construct programmable SoCs to amortize the cost of a single design across many related applications. Furthermore, the cost of errors in the design can be significant. No matter how seamless the SoC design flow is made, and no matter how careful an SoC designer is, there will inevitably be some chips that have problems that are found after fabrication. This may be due to design errors not detected by simulation or it may be due to a change in design requirements. While this type of problem is not unique to chips designed using the SoC methodology, it lends itself to the use of an elegant solution to the problem: one or more programmable logic cores can be incorporated into the SoC.

A programmable logic core (PLC) is a flexible logic fabric that can be customized to implement any digital circuit after fabrication. Before fabrication, the designer embeds a programmable fabric, consisting of many uncommitted gates and programmable interconnects between the gates, onto the chip. After the fabrication, the designer can then program these gates and the connections between them to serve different applications or implement design changes. These configurable logic blocks and connections have also been commonly referred to as embedded FPGAs (field programmable gate arrays), as opposed to stand-alone FPGAs that have been available for two decades.

Several companies already provide programmable logic cores [1]–[4]. Yet, the use of these cores is still far from mainstream. There are a number of reasons for this:

1) Tools for the design and integration of programmable fabrics are not widely available as yet. This is somewhat of a chicken-and-egg problem: existing tools and flows will not be enhanced to support the seamless integration of programmable logic cores until this design technique becomes mainstream, and the design technique will not become mainstream until the tools are enhanced to support programmable logic cores. However, as chip design costs escalate, the economics of chip design will be a strong driver for increased hardware programmability.
2) Programmable logic cores come in relatively fixed formats. That is, the integrated circuit designer can not modify the overall size of the fabric or the internal structure of the programmable logic core. The integrated circuit designer must choose a programmable logic core

Manuscript received April 1, 2004; revised August 20, 2004. This work was supported by Micronet, Altera, and a number of grants from the Natural Sciences and Engineering Research Council of Canada (including Grant STPGP 257684).

S. J. E. Wilton, J. C. H. Wu, V. O. Aken'Ova, and R. Saleh are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

N. Kafafi is with PMC-Sierra, Burnaby, BC V5A 4V7, Canada.

K. A. Bozman is with Altera, Toronto, ON M5S 1S4, Canada.

Digital Object Identifier 10.1109/JSSC.2004.841038

that is closest to the desired size; this could lead to wastage of chip area. This can be addressed by providing tiles of programmable logic that can be snapped together to form a design logic fabric of the desired size to minimize the area penalty.

3) Embedded programmable logic is not as efficient as hardwired logic in terms of area, power and speed. There are, however, special-purpose fabric generators emerging that can provide a better tradeoff between these specifications, depending on the target application.

In spite of these barriers, we believe that the use of embedded programmable fabrics will continue to increase on both ASIC and SoC designs. There will be a need for large-grain, medium-grain and fine-grain fabrics to serve a variety of needs on the chip. Of particular interest in this paper is the use of fine-grain programmable fabrics. There are many cases where an integrated circuit designer would prefer to have many very small regions of programmable logic, rather than a single or handful of large programmable logic regions. As a simple example, consider a control logic block which coordinates the operation of the rest of the chip; it may be beneficial to map selected parts of this control logic to programmable logic, rather than the entire control logic block.

In this paper, we describe a novel method for incorporating fine-grain programmable logic cores into an SoC. Rather than providing "hard" rectangular layouts, core vendors would provide "soft" descriptions of their programmable logic cores (PLC). Alternatively, the user could develop these cores themselves without much difficulty. These descriptions would typically be written at the register transfer level (RTL) in a hardware description language (HDL), such as VHDL or Verilog. We refer to this as a soft PLC. The integrated circuit designer could then incorporate the soft PLC description into the RTL description for the rest of the (nonprogrammable) chip, and then synthesize the entire chip using existing synthesis tools. The advantages and certain limitations of this approach are the subject of this paper.

In [5], Phillips and Hauck describe the Totem architecture, which is a coarse-grained programmable logic fabric. Phillips and Hauck describe several ways of implementing their fabric, one of which is to use a soft description mapped to standard cells. Unlike our approach, however, they focus on large coarse-grained fabrics rather than the small fabrics that might be incorporated into an SoC. Reference [6] also describes a standard-cell implementation of a programmable logic fabric, but again, it does not specifically target the SoC domain.

This paper is organized as follows. First, the soft PLC technique is described in more detail in Section II. Sections III and IV describes new architectures and place-and-route algorithms for these cores. Since the soft cores are intended to be synthesized using standard synthesis tools, it is unlikely that traditional FPGA architectures, optimized for full-custom layout, will be appropriate. We provide two novel architectures [7], [8] that are designed specifically for these soft cores. Section V identifies key parameters for our architectures, and seeks optimum values for these parameters. Finally, Section VI describes our experiences with a test chip that was fabricated using one of our synthesizable programmable logic cores. Conclusions are provided in Section VII.

II. Soft PLC Design Flow

As described in the introduction, integrated circuit designers who wish to use a programmable logic core typically receive a "hard core" which contains the actual physical transistor layout information. The size and shape of the core is fixed; the only freedom the designer has is where to position the core on the chip and how to connect the I/O to the block. However, using our scheme, the designer receives the core in the form of a "soft core". A "soft core" is one in which the designer obtains an RTL description of the behavior of the core, written in Verilog or VHDL. In this sense, it is similar to the definition of a soft IP core used in SoC designs [15]. The distinction is that, in a soft PLC, the user circuit to be implemented in the core is programmed after fabrication.

The value of this approach is derived from the tools needed to implement the fabric. Since the designer receives only an RTL description of the behavior of the core, synthesis tools must be used to map the behavior to gates and eventually to layout. These tools can be the same ones that are used in the standard ASIC flow. In fact, the primary advantage of the new method is that existing ASIC tools can be used to implement the chip. No modifications to the tools are required, and the flow follows a standard integrated circuit design flow. This will significantly reduce the design time of chips containing these cores.

A second advantage is that this technique allows small blocks of programmable logic to be positioned very close to the fixed logic that connects to the programmable logic to improve routability and shorten wire lengths. The use of a "hard core" requires that all the programmable logic be grouped into a small number of relatively large blocks. A third advantage is that the new technique allows users to customize the programmable logic core to better support the target application. This is because the description of the behavior of the programmable logic core is an RTL description that can be understood and edited by the user. Finally, it is easy to migrate the programmable block to new technologies; new programmable logic cores from the core vendors are not required for each technology node [15].

Of course, the main disadvantage of the proposed technique is that the area, power, and speed overhead will be significantly increased, compared to implementing programmable logic using a hard core. Thus, for large amounts of circuitry, this technique would not be suitable. It only makes sense if the amount of programmable logic required is small. In Section V, we will quantify this tradeoff, but first we explore the issues of design flow and architecture suitable for such an approach.

The basic design flow employing soft PLCs is as follows:

1) The integrated circuit designer partitions the design into functions that will be implemented using fixed logic and programmable logic, and describes the fixed functions using a hardware description language. At this stage, the designer must determine the size of the largest function that will be supported by the core; this can be done either by considering example configurations, or based on the experience of the designer.

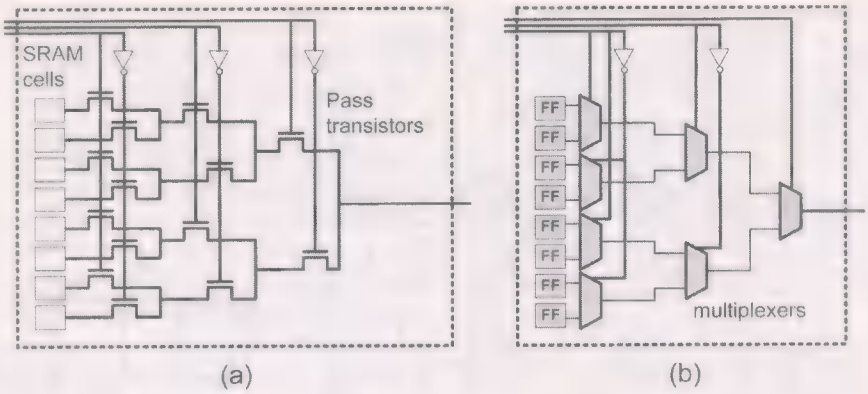


Fig. 1. Comparison of standard FPGA and soft PLC blocks. (a) standard FPGA logic block. (b) Soft PLC logic block.

2) The designer obtains an RTL description of the behavior of a programmable logic core. This behavior is also specified in the same hardware description language.
3) The designer merges the behavioral description of the fixed part of the integrated circuit (from step 1) and the behavioral description of the programmable logic core (from step 2), creating a behavioral description of the block.
4) Standard ASIC synthesis, place, and route tools are then used to implement the soft PLC behavioral description from step 3. In this way, both the programmable logic core and fixed logic are implemented simultaneously.
5) The integrated circuit is fabricated.
6) The user configures the programmable logic core for the target application.

Note that in Step 4 of the design flow, there is an important difference in the implementation of the programmable logic for a standard FPGA fabric and a soft PLC fabric, as illustrated in Fig. 1. Consider the simplified view of a 3-input lookup table (3-LUT) used in an FPGA. The standard fabric uses SRAM cells to store configuration bits and pass transistors to implement the 3-LUT shown in Fig. 1(a). In the soft PLC case shown in Fig. 1(b), a standard-cell library is used to implement the same 3-LUT. In fact, all desired functions of the soft PLC are constructed from NANDs, NORs, inverters, flip-flops (FF) and multiplexers from the standard cell library. The same holds true for the programmable interconnect in the FPGA and soft PLC.

To emphasize this point further, consider how the complete fabric would be constructed in the two cases. For the soft PLC, the final logic schematic and layout is determined by the logic synthesis tool, technology mapping algorithms, and the place-and-route tool. In the case of a hard fabric, a custom layout approach is used to create a "tile" for the FPGA. Then the FPGA fabric is assembled by replicating the tiles horizontally and vertically. Clearly, the standard FPGA approach is more area efficient but the soft PLC has the advantage of ease of use.

III. Proposed Architectures for Soft PLC

Now that the main features of the approach have been outlined, we describe two alternative architectures for a soft programmable logic core. The first proposed architecture is very similar to a standard FPGA architecture with some adjustments. However, this approach still has a significant area penalty. Since the desired fabric is intended for fine-grain programmability, one would expect the architecture to be different from standard FPGAs. As will be shown in Section V, we can reduce the area of our core by removing some degree of flexibility; the second architecture contains fewer programmable switches and hence is more area-efficient, yet contains enough flexibility to implement small circuits.

A. Architecture 1: Directional Architecture

The most straightforward way to implement a synthesizable programmable logic core is to describe the behavior of a standard FPGA at the RTL level using a hardware description language. The standard FPGA blocks are fairly complex and allow for both combinational and sequential elements. It is important to carefully consider the target applications and the required complexity of the programmable blocks. In doing so, we can make the following observations.

Observation 1: Synthesizable programmable logic cores only make sense for very small amounts of programmable logic. An envisaged application would be the next state logic in a state machine. In that case, only combinational functions are needed.

Observation 2: Many CAD tools (the tools that will be used to synthesize the programmable logic core, perform timing verification, etc.) have problems with combinational loops.

These observations motivate us to modify a standard FPGA architecture. First consider Observation 1. Since we are targeting small amounts of logic, we began with an architecture that will only implement combinational logic, allowing us to remove all flip-flops needed for sequential logic functions. Flip-flops can be added at the inputs and outputs of the programmable logic core by the IC designer if desired. Removing flip-flops reduces area and simplifies timing analysis. Of course, the flip-flops associated with the programming cells are still required for both logic and interconnect blocks.

Observation 2 leads to a more interesting problem since an un-programmed PLC contains many combinational loops. Although these loops are ultimately false paths, they can still pose problems for CAD tools and during the actual configuration bit programming process. Thus, we have created a "directional" architecture in which the flow between logic blocks can only occur

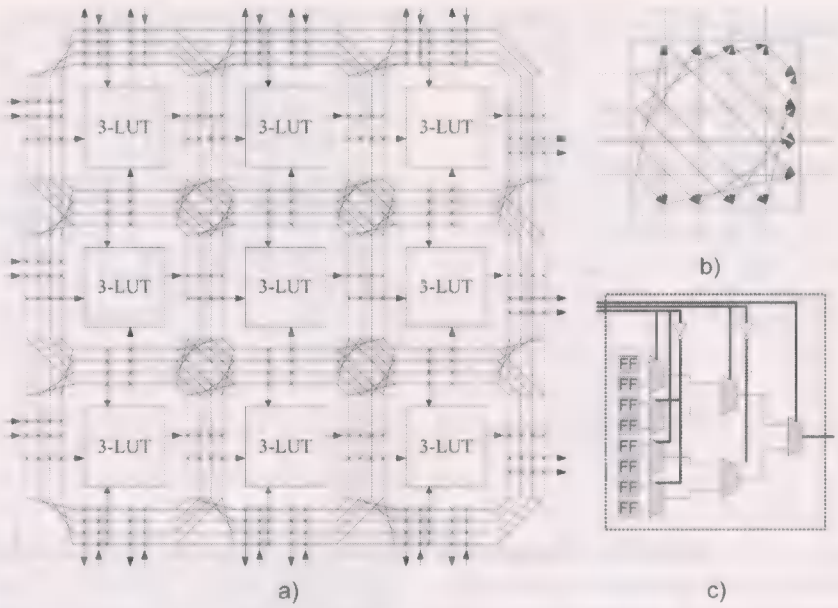


Fig. 2. Directional architecture. (a) Directional architecture. (b) Closeup of switch block. (c) 3-input look-up table (3-LUT).

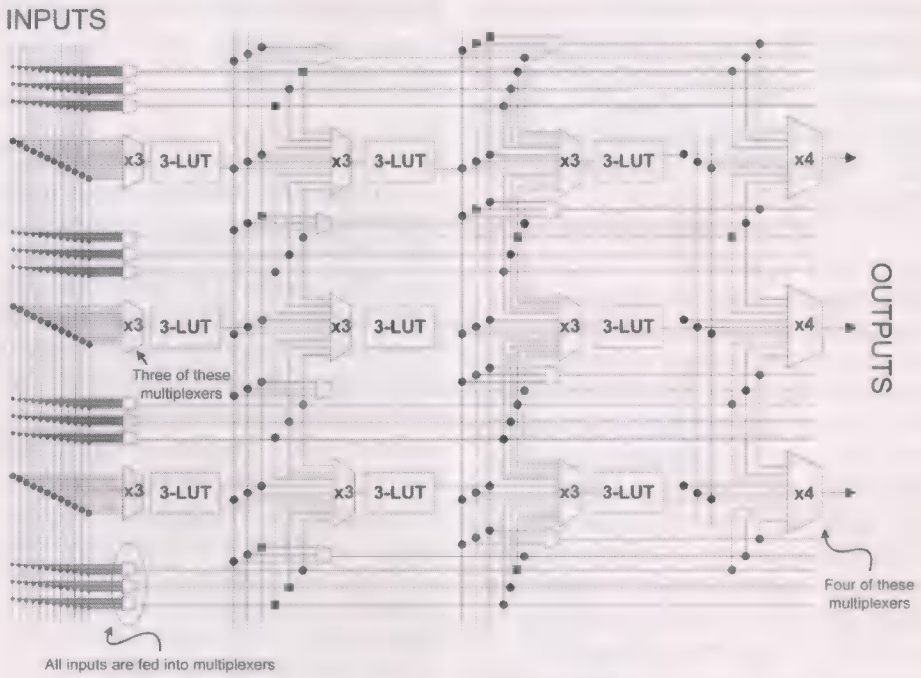


Fig. 3. Gradual Architecture.

from left to right. Since our architecture only implements combinational circuits, this will not allow any loops in the logic; any feedback loops that are required would be implemented outside of the core.

Based on these observations, we have created the architecture shown in Fig. 2(a). Each switch block is a standard switch block, with all right-to-left connections removed, as shown in Fig. 2(b). A simplified view of the 3-LUT is shown again in Fig. 2(c). The choice of a 3-LUT (as opposed to a 4-LUT or 5-LUT) was based on the observation that the ratio of logic area divided by routing area is larger in a synthesized core than a hand-optimized core; thus, we found that a smaller LUT is more efficient.

B. Architecture 2: Gradual Architecture

We can consider more efficient architectures by making the following additional observations.

Observation 3: Since we are implementing such small circuits, we should consider removing some flexibility to improve area efficiency.

Observation 4: Since the core will be hardwired into a fixed-function chip, we will require additional flexibility on the inputs and outputs.

Observation 5: Unlike a hard FPGA layout, it is not critical that each tile be identical. In a hard layout, FPGA vendors do

not wish to layout multiple tiles; in our case, the fabric is synthesized and laid out automatically by CAD tools. Therefore, we have some freedom in defining the structure of the underlying fabric.

These observations lead to the architecture in Fig. 3, which we call the "Gradual Architecture." Like the Directional Architecture, signals in the Gradual Architecture flow from left to right, and the logic resources consist only of 3-LUTs. However, in this architecture, the number of horizontal routing channels gradually increases from left to right, since more outputs are generated in each level that can be used as inputs by the downstream LUTs. The vertical tracks are only accessible through LUT outputs (each vertical track can be driven by one LUT), and can be connected to horizontal tracks using a dedicated multiplexer at each grid point. Note that, except for this multiplexer, no switch block is required in this architecture. The extension of this architecture to any number of rows and columns is straightforward.

The routing multiplexers in the first column are different from the others. We have performed experiments showing that primary inputs are frequently required in many different columns. Thus, we have included several routing multiplexers in each row (we will vary the number of these multiplexers in Section V). For each row there are one or more output select multiplexers to choose a primary output of the circuit. The output multiplexers choose between the outputs of all LUTs located in the last column and any horizontal line located above or below that specific row. The exception to this is that only one routing multiplexer per row from the first column passes a signal to the output select multiplexers.

IV. Placement and Routing Issues

Once a programmable logic core has been embedded into a chip design, and the chip has been manufactured, the user-defined circuit can be implemented on the core. A CAD tool is usually employed to determine the programming bits needed to implement the user-defined circuit. Since our architectures contain novel routing structures, some modifications must be made to standard FPGA placement and routing algorithms. In this section, we describe these modifications for the two architectures described in Section III.

It is important to note that we are not referring to the standard cell placement and routing tools needed to implement the programmable fabric itself onto the chip. Rather, the algorithms in this section are used to implement a user circuit on the programmable fabric *after the chip has been fabricated*. For example, the VPR tool [9] determines where to place the logic functions and how to form the connections between the logic functions on a given FPGA fabric. At the end of the process, the programming bits are generated for the fabric. These bits must be shifted into the fabricated chip to implement a user-defined circuit. The process is repeated if a different user circuit is to be implemented.

A. Placement Algorithms

1) Directional Architecture: The placement algorithm for the Directional Architecture described in Section III is based on the original simulated annealing placement algorithm of VPR [9]. The only change was to impose a restriction on the placer which stipulates that *input sources for all blocks must originate from the left of that block*. Otherwise, it is viewed as an illegal placement. During the annealing, we never allow a move that would result in an illegal placement.

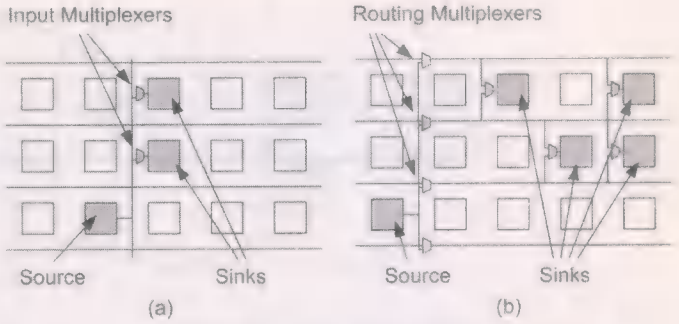


Fig. 4. Good placements on the Gradual Architecture.

The cost function used in the VPR placement algorithm depends on the delay of potential connections as well as on the Manhattan distance between pins. In a synthesized core, the delay between pins depends on where the individual cells that make up the core are positioned; it may be that adjacent blocks in the conceptual representation of Fig. 2(a) may be positioned far apart in the actual layout. However, for convenience, we base our placement cost function on the distances and delays in the conceptual representation. Improvements can be made by supplying the VPR tool with the extracted delay and distance information from the actual layout of the synthesized core. Instead of relying on the conceptual representation, we can then use the "physical" representation to obtain better delay estimates during placement and routing.

2) Gradual Architecture: In the Gradual Architecture, the routing fabric is less flexible than a standard FPGA. Poor placements can easily lead to un-routable implementations. We use a simulated annealing based algorithm with a unique cost function for this architecture, as described below.

Fig. 4 shows two examples of "good" placements on a simplified view of the Gradual architecture. In Fig. 4(a), a source logic block drives two sink logic blocks in the adjacent column. The corresponding net can be routed without any conflicts since no shared resources are required. Note that the input multiplexer used to feed each input pin of a logic block is not a shared resource; there is one such multiplexer per input pin. Any number of sinks in the column immediately adjacent to the source can be connected in this way as shown in Fig. 4(a) for the case of two sinks.

On the other hand, nets that drive logic blocks that are not in the immediately adjacent column must make use of routing multiplexers; these are shared resources. In the example of Fig. 4(b), a net drives four sinks but only needs one routing multiplexer, since the sinks are all in two vertically adjacent rows (meaning that the track between the two rows can be used to drive all sinks). If another net also required the shaded routing multiplexer, a conflict would arise when we tried to route the two nets. Since these routing multiplexers are shared resources, we wish to minimize the number of routing multiplexers used by

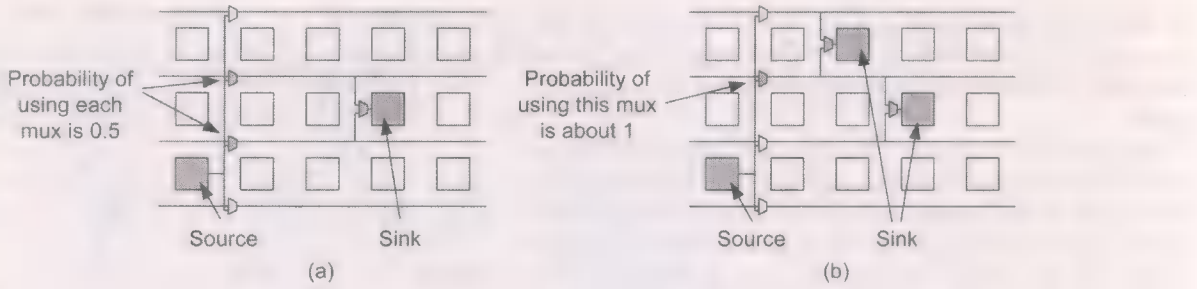


Fig. 5. Example placements on the Gradual Architecture.

each net. Therefore, we should penalize placements that generate many such potential conflicts for the router. Again note that the input multiplexers used to feed the input pins of each logic block are not shared resources, and thus should not play a role in the cost of a given placement.

Based on these considerations, a new cost function was developed for the placement algorithm that directly relates to overuse of routing multiplexers. Before presenting the cost function itself, we first describe certain factors that will be used in the function. Consider the nets in Fig. 5(a) that would connect the indicated source and sink. In this case, we consider it equally likely that the final routed net will use one of the two indicated routing multiplexers; therefore, we define the *demand* for each of the two multiplexers as 0.5 relative to the indicated source and sink. In Fig. 5(b), it is almost certain that the routed net will use the indicated routing multiplexer, since that single multiplexer can be used to feed both sinks, so the demand for that net is close to 1. Note that a valid route could be found that does not use this multiplexer; however, such a route would require two routing multiplexers. During placement, we assume that this will not happen, and thus, set the demand term for all other routing multiplexers for this net to 0. Of course, this does not mean the router is constrained to use this routing multiplexer. It is simply an assumption made to compute the cost function during placement.

Fig. 6 shows a net that drives four vertically adjacent rows. In this case, we assume that the two indicated routing multiplexers are used with probability 1 during placement. Experimentally, we have determined that this leads to better results than if we assign all five routing multiplexers in that column the same value (which would be about 1/2). Again, note that the router is not constrained to actually use the indicated multiplexers.

To derive the cost function, we start by defining an occupancy function, $\mathrm{Occ}()$, of a routing multiplexer as an estimate of how many nets would like to use that routing multiplexer. We can write this as the sum of the estimated demand for a given multiplexer by each net:

$$\mathrm{Occ}(c,r) = \sum_{n \in \mathrm{Nets}} \mathrm{demand}(c,r,n)$$

where $\mathrm{demand}(c,r,n)$ is the estimated demand for the routing multiplexer at column and row (c, r) by net n. As already described, the demand is a number lies in the range between 0 and 1; 0 implies that there is little chance that the router will use this multiplexer to route net n, while 1 means that the router will, with high probability, use this multiplexer when routing net n.

Next we define the capacity function, $\mathrm{Cap}()$, of a routing multiplexer as the number of output lines available from a given

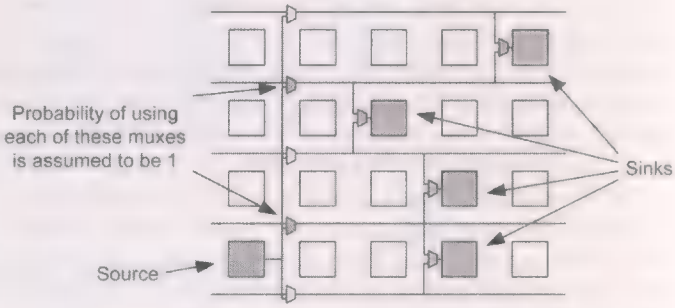


Fig. 6. Example placements on the Gradual Architecture: Sinks in many adjacent rows.

set of input lines. It is an estimate of the ability to satisfy the routing demand at a given location. Typically, the capacity of all routing multiplexers is set to 1 since each one has a single output. However, for those muxes in the first column, the capacity is equal to the number of horizontal lines that can be driven from primary inputs. Referring back to Fig. 3, the capacity function would be 3 since three muxes drive 3 adjacent horizontal lines from the same set of primary inputs at each location.

With these definitions in place, the cost of a given placement on a C-column, R-row core is given by

$$\mathrm{Cost} = \sum_{r=0}^{R} \sum_{c=0}^{C} \max[0, (\mathrm{Occ}(c,r) - \mathrm{Cap}(c,r) + \gamma)]$$

where $\mathrm{Occ}(c,r)$ is the occupancy demand of a routing multiplexer at location (c,r), and $\mathrm{Cap}(c,r)$ is the output capacity of multiplexers at location (c,r). We take the difference between $\mathrm{Occ}()$ and $\mathrm{Cap}()$ to incorporate the fact that one or more outputs are available at each location. If the difference is negative, we set the cost of that routing mux to 0 using the max function. The γ term is a small bias value (set to 0.2 for our experiments).

B. Routing Algorithms

The negotiated-congestion based routing algorithm from VPR [9] was used without modification for both architectures. For the Gradual Architecture, the routing task is very easy since there are only a few potential routes for each net. For the Directional Architecture, there are many potential routes so the routing is more complex. The use of the advanced router within VPR gave us ability to evaluate different architectures and placement schemes during our architectural investigation.

TABLE I
DIRECTIONAL AND GRADUAL ARCHITECTURE RESULTS

Benchmark Circuit	Directional Architecture			Gradual Architecture		
	FPGA Core Size	Tracks per Channel	Cell Area (μm^2)	FPGA Core Size	Input Muxes per row	Cell Area (μm^2)
cc	9x9	4	300 460	8x8	3	263 101
cm138a	5x5	3	80 868	5x5	1	78 319
cm150a	9x9	4	300 460	8x8	3	263 101
cm151a	5x5	3	80 868	4x4	1	43 932
cm152a	4x4	3	53 004	4x4	1	43 932
cm162a	5x5	4	96 854	5x5	2	89 614
cm163a	6x6	5	174 589	5x5	2	89 614
cm42a	5x5	4	96 854	5x5	2	89 614
cm82a	4x4	3	53 004	2x2	1	9 261
cm85a	6x6	4	137 518	6x6	2	128 822
cmb	7x7	3	154 407	7x7	2	184 590
comp	12x12	4	528 332	11x11	2	542 489
con1	4x4	3	53 004	4x4	1	43 932
count	12x12	5	667 344	10x10	4	487 588
cu	8x8	3	199 702	8x8	2	244 676
5xp1	11x11	5	562 305	11x11	2	542 489
i1	8x8	3	199 702	7x7	2	184 590
inc	10x10	5	466 121	10x10	2	424 445
unreg	10x10	4	368 620	9x9	4	388 074
Average			240 737			218 009
Geo. Avg.			**175 014**			**141 954**

V. EXPERIMENTAL RESULTS

In this section, we experimentally compare the two architectures described in Section III. We used 19 small combinational MCNC benchmark circuits [14]. We selected small circuits since these are the type of circuits we expect to be used with our architecture; large circuits would likely be implemented using hard programmable logic cores. For each circuit, we initially found the minimum-size square core on which the circuit can be placed and routed. We then created a VHDL description of each core, and synthesized it using Synopsys Design Compiler™ and a standard 0.18-μm CMOS library. The cell area reported by the Synopsys tool was used for a basis for comparison in Table I.

A. Directional Architecture Versus Gradual Architecture

The first four columns of Table I show the results for the Directional Architecture. For each benchmark circuit, we varied both the core size and the number of tracks in each channel, and chose the configuration which resulted in the minimum area; the chosen size and channel width are shown in columns two and three of the table. For each configuration, we then synthesized the architecture using Synopsys; the fourth column in the table shows the cell area required to implement the core.

The final three columns show the results for the Gradual Architecture. In this case, we varied both the core size and the number of input multiplexers per row, and chose the configuration which resulted in the lowest area. These numbers are reported in columns five and six of the table, and the synthesized cell area from Synopsys is shown in the final column. From the last row of the table, the geometric average of the area required to implement the circuits on the Gradual Architecture is 18.9% less than that required to implement the same circuits using the Directional Architecture.

B. Soft Versus Hard Programmable Logic Cores

As mentioned in Section II, the primary disadvantage of using a "soft" programmable logic core is the reduced density, speed, and increased power consumption. In this subsection, we estimate the area penalty of a soft core compared to a hard core.

The most accurate way to compare the area required by soft and hard programmable logic cores would be to lay out (by hand) a hard core, and compare its area with the numbers in Table I. This is a time-consuming task. Instead, we estimated the size of a hard core using a detailed transistor-count model, following the methodology described in [9]. We focus on a 4×4 Gradual Architecture with three input multiplexers per row. By estimating the number of minimum transistor equivalents (MTEs) required to implement the circuit, and converting this to area in our 0.18-μm technology, we estimate the layout area of such a core to be 12868 μm^2. A soft core was generated using these same parameters, and the size (after synthesis using Synopsis and physical design using Cadence) was 81092 μm^2. Thus, the synthesized core requires approximately 6.4× more area than the hard core.

This number is significant. Clearly, for large programmable logic cores, our approach would not be suitable. However, if only small amounts of programmable logic are required, this density penalty may be acceptable. In addition, the use of a hard core will usually require the selection of a core from a library. Since it is unlikely that a library would contain all sizes and shapes of cores, in most cases, a designer would end up choosing a larger core than is required. Using a soft core, the designer can create a core of any size. Even if a core of the appropriate size was created, the difficulty inherent in embedding hard cores may make the use of hard cores less attractive than our soft approach.

We have also compared our sizes to commercial FPGA layouts using publicly available information. These comparisons

TABLE II
SENSITIVITY OF RESULTS

I/O Connections	Grad vs. Dir
Default I/O connections	18.9 %
Half as many I/O connections	9.67 %
Twice as many I/O connections	2.33 %
Margin	9.23 %
Conclusion	Sensitive

Simulated Annealing Algorithm	Grad vs. Dir
Percent Difference, baseline algorithm	18.9 %
Percent Difference, fast algorithm	15.5 %
Margin	3.4 %
Conclusion	Slightly Sensitive

yield little insight, however, since the commercial devices contain far more tracks per channel, and contain additional elements such as flip-flops in the logic blocks.

C. Sensitivity of Results

As described in [11], it is critical to analyze results for their sensitivity to experimental assumptions. Table II shows two of our sensitivity results for the data in Table I. The first part of the table shows how the conclusions change if we alter the number of input/output connections per grid. In the experiments in Section V-A, it was assumed that an $n \times n$ Directional Architecture has $2n$ input/output connections along each of the four edges of the core, and that an $n \times n$ Gradual Architecture has $4n$ input/output connections along the left and right edges of the core. We attempted to use two other input/output ratios, and gathered the results in Table II. Although the Gradual Architecture always produced higher density than the Directional architecture, the margin by which the Gradual was better varied (we do not have enough data to conclude that this is a result of anything other than experimental "noise"). According to the methodology in [11], we classify this experiment as sensitive to the input/output ratio, even though the conclusion that Gradual is better than Directional was the same in all cases.

The second part of the table shows how a less aggressive placement schedule (fewer moves per temperature and larger temperature drops during the annealing) and routing schedule (fewer routing attempts) affects the conclusions. In this case, the margin was smaller, meaning the experiment was only slightly sensitive to the choice of algorithm.

D. Nonrectangular Fabric

The grid of logic blocks in standard FPGAs is usually square or rectangular. From [12], however, logic circuits often have a "triangular" shape as shown in Fig. 7(a). In standard FPGAs, this does not present a problem, since the routing resources are flexible enough that signals can be routed left, right, up, or down, as shown in Fig. 7(b). This means that in a standard FPGA, the physical implementation of a circuit need not match the fanout shape of the circuit. In the architectures described in this paper, however, the signal flow is restricted from left to right. As shown in Fig. 7(c), this can lead to unused logic blocks if the circuit does not have a naturally square shape.

We can alleviate this problem somewhat by creating a programmable logic core that is not square. We have observed that in many implementations, several logic blocks in the rightmost columns remain unused. We can take advantage of this by removing logic blocks from the last few columns, as indicated with shading in Fig. 7(c). We quantify the number of logic blocks removed using the parameter c, where c is defined as

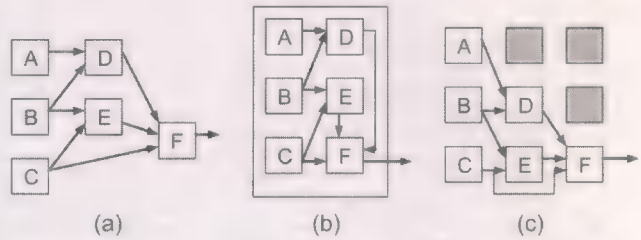


Fig. 7. Implementing a circuit on a triangular core.

the proportion of the logic blocks in the top row that have been removed. In Fig. 7(c), c is 2/3. In all cases, we remove blocks in a "triangular" fashion; if we remove m blocks from column i, we remove $m-1$ blocks from column $i-1$. A value of 0 for c indicates a rectangular core; a value of 1 indicates a triangular core. Note that a nonzero value of c does not imply a nonrectangular final layout. The diagram in Fig. 7(c) is a conceptual representation; the core will be synthesized into gates, and the gates will be placed into rows of standard cells regardless of the shape of the conceptual representation. Intuitively, as c is increased, the area of the implementation will go down. If c is decreased too much, however, the area will rise, since a larger virtual grid will be needed. This effect can be seen in Fig. 8. Fig. 8(a) shows how the implementation area depends on c for each circuit implemented on the Gradual Architecture (each line represents a different circuit). Because we were unable to synthesize large triangular cores using our synthesis tools, results are only shown for 11 of the 19 benchmark circuits. The geometric average over these 11 circuits is shown in Fig. 8(b).

Although each individual circuit in Fig. 8(a) exhibits its own characteristics, the results in Fig. 8(b) indicate that the overall gain obtained using a nonzero value of c is relatively small. From Fig. 8(a), the "breakpoint" (the point at which a larger grid is needed) is not the same for each circuit. Thus, the average results show that only a modest improvement can be achieved. Overall, the value of c that gave the lowest area was 0.6, which resulted in an 11.1% lower area than a square core, averaged over all circuits.

VI. PROOF-OF-CONCEPT IMPLEMENTATION

To investigate the implementation issues of our synthesizable embedded core approach, we have chosen a module derived from a chip testing application. This module acts as a bridge between a test access mechanism (TAM) circuit [13] and an IP core under test. In the research work described in [13], the TAM is actually a communication network that transfers test data to/from internal IP blocks on the chip in the form of packets. The module we selected allows the TAM and the IP core to run at different frequencies, resulting in higher overall TAM throughput. A chip

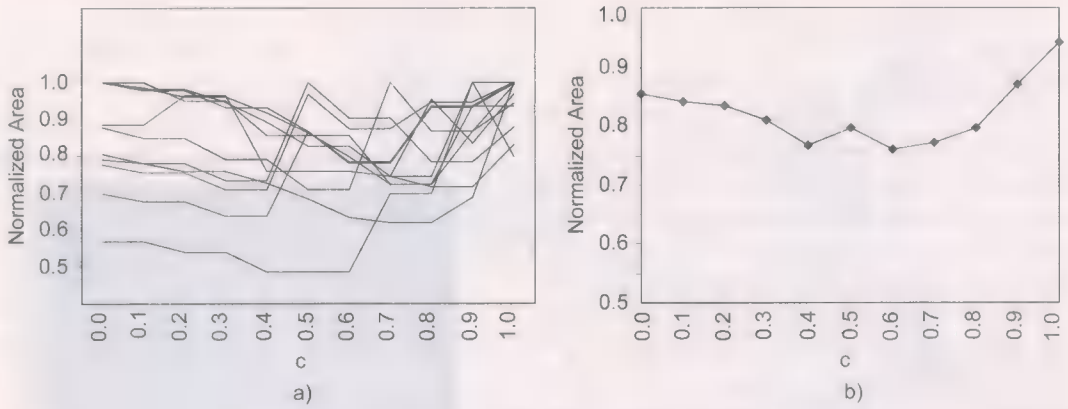


Fig. 8. Area as a function of c for Gradual Architecture. (a) One trace per benchmark circuit. (b) Geometric average over benchmark circuits.

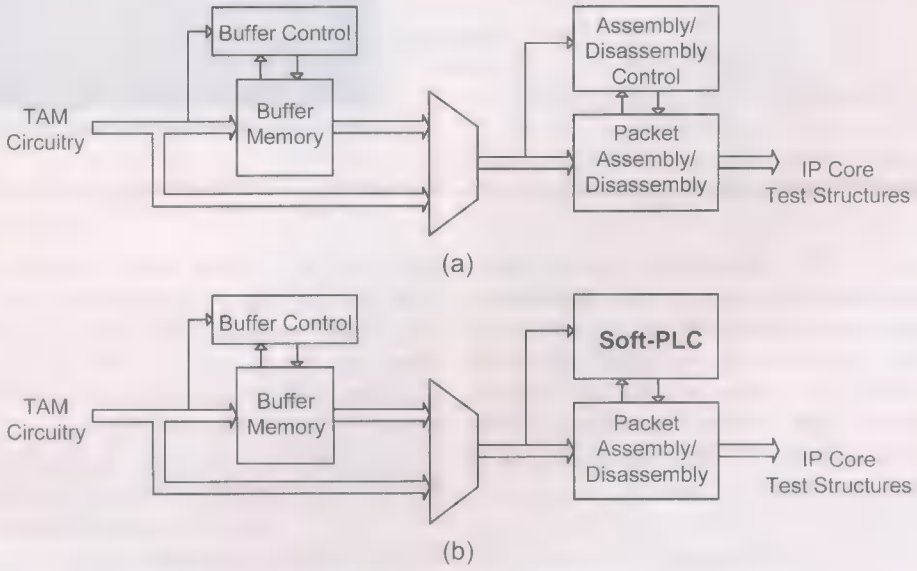


Fig. 9. Schematic of proof-of-concept module. (a) TAM-IP interface module (nonprogrammable). (b) TAM-IP interface module (programmable).

designed with this type of network TAM would contain one of these selected modules for each IP core on the chip.

A. Reference Version

Fig. 9(a) shows a block diagram of the module. The module consists of a buffer memory, a packet assembly/disassembly block, and two state machines. Packets received from the TAM circuit are optionally buffered before being converted to a form usable by an IP core under test. A key component in the module is the Packet Assembly/Disassembly block which controls the assembly and disassembly of test packets based on a given packet format. The packet format was subject to change from time to time during the course of the research described in [13] which required a re-design of this block.

B. Programmable Version

When packet formats are modified to adjust header, data and address information, the control circuitry must also be modified. Noting this fact, we decided that the next-state logic would benefit from programmability. This would allow the user to modify some packet processing and control operations simply by re-programming the block. If the next state logic of the state machine is made programmable, as shown Fig. 9(b), new schemes can be implemented after fabrication of the integrated circuit. Although a hard programmable logic core could also be used here, it is better suited to the soft PLC approach due to its fine-grain nature.

C. Implementation Issues

We designed two versions of this module: 1) the reference version with no configurability, and 2) the programmable version, in which the assembly/disassembly control is removed and replaced with a soft programmable logic fabric. The fabric uses the Gradual Architecture as it was found to be more efficient than the Directional Architecture. When adding the programmable component to our module, a number of other interesting issues arose. This section summarizes these issues.

1) Programmable Logic Core Size: The first issue was how much programmable logic is needed to replace the fixed next state logic. Without knowing the actual logic function that will eventually be implemented in the core, it is difficult to estimate the amount of programmable logic required. However, in this case, we have domain knowledge regarding the types of functions that will be implemented, and we can use this knowledge

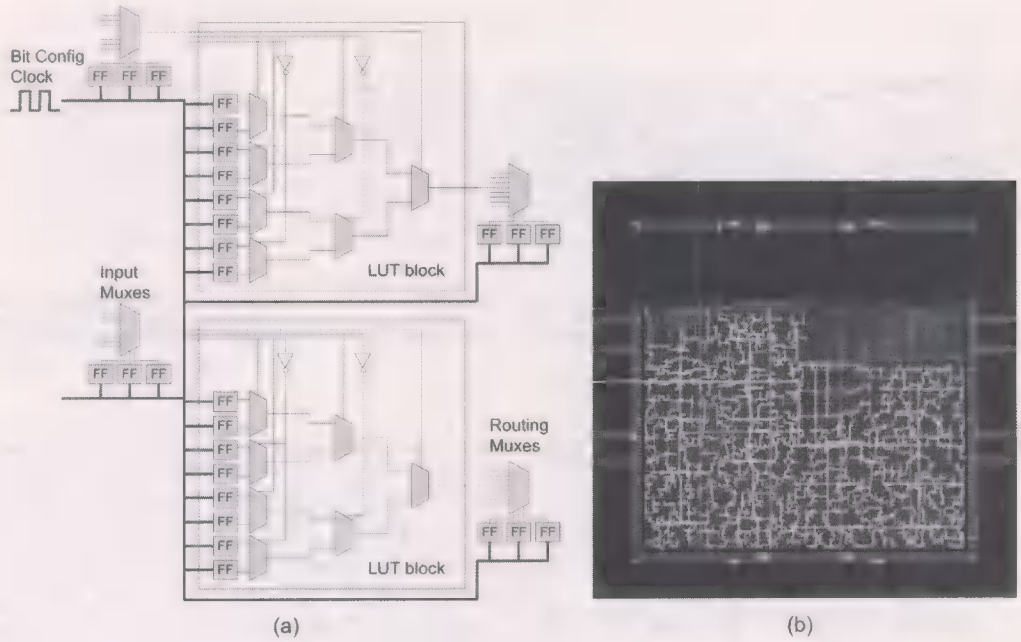


Fig. 10. Programming clock tree routing complexity. (a) Portion of Gradual Architecture. (b) Physical design of programmable module.

to make reasonable decisions. We designed two user logic functions that would be implemented in the core, and determined the size of the core that would be required to implement each function using VPR [9]. For our circuit, we found that a core consisting of 49 LUTs (i.e., a 7×7 array of 3-LUTs) would be sufficient for both potential logic functions; however, to allow some safety margin and anticipation of larger functions, a core of 64 LUTs (8×8 array) was used.

2) Connections Between the Core and the Fixed Logic: A second issue is how the programmable logic core is connected to the rest of the module. Although the core itself is programmable, specific inputs and outputs must be connected to the core in advance. This will dictate which functions are possible to implement in the core. Again, we have domain knowledge to assist us with this decision. We can select which inputs are connected to the core and which outputs will be made available from the core. In our design, the two user logic functions required 9 inputs and 10 inputs, respectively, and required 11 outputs and 12 outputs, respectively. We afforded ourselves some flexibility by hardwiring a selected set of 10 inputs and 13 outputs to our core.

3) Routing the Programming Clock Signal: During physical design process, it was apparent that our synthesizable core was placing an extra burden on the router due to the large number of flip-flops in the design. A programmable logic core contains many configuration bits to store the state of individual routing switches and the contents of lookup tables; in a synthesizable core, these configuration bits are built using flip-flops that have clock inputs to enable programming. As shown in Fig. 10(a), there are configuration bits for input muxes and output muxes, as well as the LUTs themselves. Each of these FFs must be connected to a common clock signal for programming purposes as indicated by the bold line.

To determine how flip-flop-intensive our core is, we compared its flip-flop density to that of a nonprogrammable design. We analyzed an ASIC implementation of a 68HC11 core, and found that the flip-flop density (number of flip-flops per unit area) was $1/3$ of the flip-flop density in our programmable logic core. Thus, we realized that the clock tree in our core will be more complex and consume more chip area than a typical ASIC. This was confirmed; in our implementation, 45% of the layout area was consumed by the clock tree, power striping, and signal routing (experience with other ASICs of this size has shown that 25% is usually enough). Furthermore, FFs must be connected as one long shift register for programming purposes, and this also added to the routing complexity.

The results of the physical layout of the bit configuration clock routing are shown in Fig. 10(b). Our core contains 1803 such flip-flops, each connected to the bit configuration clock signal. The clock net highlighted in white is the configuration clock; this routing is clearly more complex than the other nets (shown in grey). This extra clock complexity increases the area overhead of the design, beyond what would be estimated by just considering only the standard cell area. In our case, this is a notable source of area overhead, since the original next state logic was purely combinational logic with no FFs or clocks. Note that this clock tree overhead would occur in both a soft and hard programmable logic core.

D. Implementation Results

1) Area Overhead: We implemented both the programmable and nonprogrammable versions of the module using the same tool flow to further quantify the area overhead. The reference module (without the programmable logic core) required 369 700 μm^2 in a 0.18-μm TSMC process, of which 1217 μm^2 is the area due to the assembly/disassembly controller next state logic. The programmable module (containing 64 LUTs as described above) required 1 025 000 μm^2, of which 684 600 μm^2 was due to the programmable next state logic.

The layout areas are summarized in Table III. Clearly, the differences in these numbers are significant. Our synthesizable

TABLE III
AREA RESULTS SUMMARY

Implementation Method	Area of Next State Logic	Area of Entire Chip
Non-Programmable (measured)	1 217 μm^2	369 700 μm^2
Hard Prog. Core (estimated using results from [9])	107 000 μm^2	481 600 μm^2
Synthesizable Core (measured)	684 600 μm^2	1 025 000 μm^2

TABLE IV
SPEED RESULTS

Implementation Method	Critical Path of Module (using first user-defined logic function)	Critical Path of Integrated Circuit (using second user-defined logic function)
Reference Module (no programmability)	25.40 ns	25.40 ns
Programmable Module (with synthesizable core)	51.08 ns	51.08 ns

programmable logic core required 560× more chip area than the fixed logic that it replaced. From the analysis in Section V, the synthesizable core requires 6.4× more area than a hard programmable logic core. However, the use of a hard core may not be suitable for such fine-grain applications. It would require the same considerations as any other hard IP plus additional ones for programmability. For the size of fabric being used, the soft PLC would provide a more seamless approach.

Further investigation into the area overhead showed that 53% of the area of our programmable logic core was due to routing multiplexers and the configuration bits that control these multiplexers, as shown in Fig. 10(a). These multiplexers are large; the largest in our core has 26 inputs. Our standard cell library contains only two- and four-input multiplexer cells; larger multiplexers are built by cascading these smaller multiplexers. Clearly, the area overhead could be improved significantly by either supplementing our cell library with larger multiplexers, or modifying the architecture to employ smaller multiplexers.

2) Delay Overhead: We measured the speed of our reference and programmable modules before and after physical design. Table IV shows the post-physical design results. In this case, we configured the core using the two user-defined logic functions mentioned above, and measured the length of the critical path through the logic circuit in each case. As the table shows, the results indicate that the programmable core has approximately twice the critical path delay as the reference design, for both user-defined functions.

The module containing the programmable fabric was fabricated in 0.18-μm TSMC CMOS and tested using the same two user-defined logic functions. The speed results correlated well with the results shown above. The chip design had a critical path of about 40 ns compared to the expected 50 ns, well within the error tolerances of the models used in the CAD tools and the statistical variations of the CMOS process.

VII. CONCLUSION

In this paper, we have presented two new architectures for synthesizable programmable logic cores. Synthesizable programmable logic cores are different than the programmable cores currently available from vendors in that they are obtained as a HDL description, and synthesized using standard synthesis tools. The use of these cores has significant area overhead; we have estimated an overhead of 6.4× compared to using "hard" programmable logic cores. Yet, for small logic circuits, these "soft" cores have a number of advantages: they are easy to integrate with fixed logic, we can create cores of any size and shape, and they are easy to migrate to a new technology.

One of the primary applications we envisage for these cores is the implementation of small combinational logic blocks, such as the next-state logic or output-logic of state machines. As a result, our architectures are different than traditional FPGAs in that they only support combinational circuits, and are "directional" in that signal only flow in one direction through the fabric. In addition, the interconnect pattern is less flexible and the routing resources less plentiful. We have performed experiments to show that small combinational circuits can be implemented on these cores efficiently.

This paper also has illustrated some the issues that arise when such a core is used, through the use of a proof-of-concept chip: the choice of the size of a core, the choice of inputs and outputs, and the difficulty in routing the flip-flops.

Better synthesis results could be obtained by adding specialized cells to the standard-cell library to implement our programmable logic fabric. We have not considered this in this paper, since our goal was to create architectures that can be implemented using the standard synthesis tools, cell libraries, and design flows that are already familiar to integrated circuit designers. However, initial experiments have shown that, by removing unnecessary features, we can create a replacement for our flip-flop standard cell that is 40% the size of the standard cell version. Since, in the entire fabric, the flip-flops account for 43% of the chip area, we would expect significant savings if this standard cell was used to construct our fabric. We also expect that significant improvements can be obtained using custom-designed multiplexer standard cells. Clearly, if this design technique is to become mainstream, specialized standard cells should be created.

Although these soft cores are less efficient than their fixed counterparts, the use of programmable logic cores, and especially synthesizable programmable logic cores, is still important. The post-fabrication flexibility that these cores provide will be vital as integrated circuits get larger and as masks get more expensive. Synthesizable programmable logic cores are a sensible solution when only small amounts of programmable logic are required, since they can be treated much like regular logic during the design process. The results of this paper clearly show that there is still work to be done improving their area and speed, but as new architectures are uncovered, and new CAD techniques are developed, it is likely that both hard and soft cores will become an important part of future integrated circuits.

REFERENCES

[1] *VariCore Embedded Programmable Gate Array Core (EPGA) 0.18 μm Family*, Dec. 2001. Actel Corp., Datasheet.

[2] *HyperBlox FP Embedded FPGA Cores*, 2002. Leopard Logic Inc., Product Brief.

[3] M2000 FLEXEOStm Configurable IP Core [Online]. Available: http://www.m2000.fr

[4] eASIC 0.13 μm Core [Online]. Available: http://www.easic.com/products/easicore013.html

[5] S. Phillips and S. Hauck, "Automatic layout of domain-specific reconfigurable subsystems for systems-on-a-chip," in *Proc. ACM Int. Symp. Field-Programmable Gate Arrays*, Feb. 2002, pp. 165–176.

[6] R. Osann, S. Eltoukhy, S. Mukund, and L. Smith, "Programmable logic array embedded in mask-programmed ASIC," World Intellectual Property Org. Patent #WO 01/63766 A2, Feb. 2001.

[7] N. Kafafi, K. Bozman, and S. J. E. Wilton, "Architectures and algorithms for synthesizable embedded programmable logic cores," in *Proc. ACM Int. Symp. Field-Programmable Gate Arrays*, Monterey, CA, Feb. 2003, pp. 1–9.

[8] J. C. H. Wu, V. Aken'Ova, S. J. E. Wilton, and R. Saleh, "SoC implementation issues for synthesizable embedded programmable logic cores," in *Proc. IEEE Custom Integrated Circuits Conf.*, San Jose, CA, Sep. 2003, pp. 45–48.

[9] V. Betz and J. Rose, "VPR: A new packing, placement, and routing tool for FPGA research," in *Proc. Int. Workshop Field Programmable Logic and Applications*, Sep. 1997, pp. 213–222.

[10] V. Betz, J. Rose, and A. Marquardt, *Architecture and CAD for Deep-Submicron FPGAs*. Boston, MA: Kluwer, 1999.

[11] A. Yan, R. Cheng, and S. J. E. Wilton, "On the sensitivity of FPGA architectural conclusions to the experimental assumptions, tools, and techniques," in *Proc. ACM Int. Symp. Field-Programmable Gate Arrays*, Feb. 2002, pp. 147–156.

[12] M. Hutton, J. Rose, J. Grossman, and D. Corneil, "Characterization and parameterized generation of synthetic combinational benchmark circuits," *IEEE Trans. Computer Aided Des.*, vol. 17, no. 10, pp. 985–996, Oct. 1998.

[13] M. Nahvi and A. Ivanov, "A packet switching communication-based test access mechanism for system chips," in *Proc. IEEE Eur. Test Workshop*, 2001, pp. 81–86.

[14] S. Yang, "Logic Synthesis and Optimization Benchmarks, Version 3.0," Microelectronic Center of North Carolina, Tech. Report, 1991.

[15] M. Keating and P. Bricaud, *Reuse Methodology Manual*. Boston, MA: Kluwer, 1999.

Steven J. E. Wilton (S'86–M'97–SM'03) received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Toronto, Canada, in 1992 and 1997, respectively.

In 1997, he joined the Department of Electrical and Computer Engineering at the University of British Columbia, Canada, where he is now an Associate Professor. During 2003 and 2004, he was a Visiting Professor in the Department of Computing at Imperial College, London, U.K., and at the Interuniversity MicroElectronics Center (IMEC), Leuven, Belgium. He has also served as a consultant for Cypress Semiconductor and Altera Corporation. His research focuses on the architecture of FPGAs, and the CAD tools that target these devices.

In 2005, Dr. Wilton was the Program Chair for the ACM International Symposium on Field-Programmable Gate Arrays. He is also a member of the program committee for the IEEE Custom Integrated Circuits Conference, the International Conference on Field-Programmable Logic and Applications, and the International Conference on Field-Programmable Technology, and has served as a Guest Editor for two issues of the IEEE JOURNAL OF SOLID-STATE CIRCUITS. In 1998, he won the Douglas Colton Medal for Research Excellence for his research into FPGA memory architectures. He received Best Paper Awards at the International Conference on Field-Programmable Logic and the International Conference on Field-Programmable Technology in 2001 and 2003, respectively.

Noha Kafafi received the B.A.Sc. degree in electrical and computer engineering from the University of British Columbia, Canada, in 2002. In 2003 she received the M.A.Sc. degree in electrical and computer engineering from the University of British Columbia.

Since May 2004 she has been at PMC-Sierra, where she is involved in validating high-speed communication ICs. Her research interests include FPGA CAD algorithms and architectures as well as digital IC design and testing.

James C. H. Wu (S'98–M'04) received the B.A.Sc. degree in electrical and computer engineering from the University of British Columbia, Canada, in 2002. He is currently working toward the M.A.Sc. degree, also in electrical and computer engineering from the University of British Columbia.

His research interests include field-programmable gate arrays and integrated-circuit CAD development.

Kimberly A. Bozman received the B.Eng. degree from the University of British Columbia, Canada, in 2002. After graduation, she joined the System on a Programmable Chip group at the University of British Columbia where she researched synthesizable programmable logic IP for System on a Chip design.

Since October 2002, she has been with Altera Corporation, Canada, where she is engaged in the development of commercial place and route tools. Her research interests include field programmable gate array (FPGA) architectures and CAD tools for FPGAs.

Victor O. Aken'Ova received the B.A.Sc degree in electrical and computer engineering from the University of British Columbia, Vancouver, Canada, in 2002. He is currently working toward the M.A.Sc. degree in electrical and computer engineering at the University of British Columbia.

His research interests include circuit level design, analysis, and automation.

Resve Saleh (M'79–SM'95) received the B.S. degree in electrical engineering from Carleton University, Ottawa, Canada, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley.

He is currently the NSERC/PMC-Sierra Chairholder in the Department of Electrical and Computer Engineering at the University of British Columbia, Vancouver, Canada, working in the field of system-on-chip design, verification and test. He was a founder and former Chairman of Simplex Solutions which developed CAD software for deep-submicron digital design verification. Prior to starting Simplex, he spent nine years as a Professor in the Department of Electrical and Computer Engineering at the University of Illinois in Urbana, and one year teaching at Stanford University. Before embarking on his academic career, he worked for Mitel Corporation, Ottawa, Canada, Toshiba Corporation, Japan, Tektronix, Beaverton, OR, and Nortel, Ottawa, Canada. He has published over 50 journal articles and conference papers.

Dr. Saleh received the Presidential Young Investigator Award in 1990 from the National Science Foundation. He served as general chair (1995), conference chair (1994), and technical program chair (1993) for the IEEE Custom Integrated Circuits Conference. He recently held the positions of Technical Program Chair, Conference Chair and Vice-General Chair of the International Symposium on Quality in Electronic Design, and has served as Associate Editor of the IEEE TRANSACTIONS ON COMPUTER AIDED DESIGN. He recently co-authored a book entitled *Design and Analysis of Digital Integrated Circuit Design in Deep Submicron Technology*.

Low Standby Power State Storage for Sub-130-nm Technologies

Lawrence T. Clark, *Senior Member, IEEE*, Franco Ricci, and Manish Biyani

***Abstract*—Handheld and other battery-powered ICs require process scaling to increase functional integration and reduce active power consumption. Scaling also increases leakage current components to the point where standby power is frequently a limiting design factor. A scheme combining low-leakage thick-gate shadow latches and high-performance transistors is presented that decouples performance from standby power in sub-130-nm technologies. Circuit design and operation, including pulse-clocked latches, use of dynamic circuits, and inclusion of scan is presented. The approach is validated by experimental results on a 90-nm process.**

***Index Terms*—Leakage currents, logic circuits, low power, sequential logic circuits.**

I. Introduction

INTEGRATED circuits designed for handheld and cell phone applications must meet stringent energy requirements due to limited battery capacity. Long device idle times make standby power a limiting factor in battery lifetime. Simultaneously, lower operating voltages reduce active power by the well-known quadratic factor. Additionally, lower operating voltages are required by process scaling, which in turn, drives lower threshold voltage (V_t) to maintain gate overdrive as the power supply, V_{CC} is scaled. Unfortunately, this increases transistor sub-threshold currents exponentially, leading to tradeoffs between active and standby power in process selection unless standby leakage is reduced by circuit design. Various schemes, primarily focusing on application of reverse-body bias (RBB) [1]–[4], or MTCMOS approaches [5]–[7], have been suggested and used in products to address the primary leakage components. These are the transistor off-state drain to source leakage (I_{off}), as well as drain to bulk components, due to gate induced drain leakage (GIDL) or direct tunneling from drain to bulk in transistors with steep doping profiles, especially those with pocket or halo implants [3], [8]. Since it is costly in terms of both time and power to save and then restore the state of an IC, it is imperative that any implementation be state retentive [19].

At the 130-nm technology node and beyond, oxide scaling produces significant gate oxide leakage (I_{gate}) contribution due to direct band-to-band tunneling [9] since it must keep pace with transistor channel length to maintain adequate control [9], [10]. Consequently, leakage reduction schemes for this and future technology nodes need to address this increasingly important component. The alternative is to use a thicker oxide and sacrifice performance by attempting to make up for loss of transistor gate control by very high doping.

Low standby power is frequently achieved by limiting transistor scaling to avoid leakage increases. However, the power supply voltage (V_{CC}) reduction that scaling allows is the best method to limit active power, as illustrated in Fig. 1 comprising 0.18-μm microprocessor performance on two otherwise identical processes having V_t differing by 110 mV, equivalent to about $25\times$ I_{off} leakage reduction. The 390 mV data is calibrated to an existing design that includes a low-standby-power mode combining RBB and power supply collapse [3], [11] while the 500 mV V_t data is simulated. For each data point on the curves, the processor is run at the maximum frequency allowed for the given voltage, while in the low V_t combined with a low-standby-power mode case, excess cycles are spent in the low-standby-power state. Voltage was scaled upwards in 100 mV increments from 0.6 V as required by performance.

The lower curve in the figure shows that introduction of a RBB low-power state, time multiplexed with active operation, can simulate a lower leakage process, while retaining the higher performance and lower power at high frequencies. The zero frequency points show that identical standby currents can be obtained, while at 400 MHz, with V_t of 390 mV power is 42% lower than with V_t of 500 mV. The potentially decreasing efficacy of RBB modes in future high-performance processes [12], experience in practical application, where maintaining state in domino circuits and imbalanced latches limits voltage collapse [3], and desire to make the low-power mode operable at sub-1 V V_{CC} and hence more compatible with dynamic voltage scaling (DVS) led us to investigate alternative schemes. Regardless of the actual power savings approach employed, as long as such schemes are state retentive and invoke a small power penalty upon entrance and exit, the analysis embodied in Fig. 1 applies.

In this paper, circuits to implement the low-power state, which addresses the increasing leakage components that face sub-130-nm technologies is presented. This is accomplished by placing the IC state in latches fabricated using thick-gate, high-V_t transistors and cutting off the supply to the nonstate logic circuitry. This decouples the performance of the IC in active operation from the standby power, affording more aggressive scaling to even very power sensitive handheld devices such as cell phones and personal digital assistants.

While the experimental circuits were fabricated in a 90-nm technology, the circuits and methods are applicable to future processes beyond the 65-nm technology node. Section II addresses the basic circuit design and operation. Section III describes the use in time borrowing latches and dynamic circuits,

Manuscript received January 6, 2004, revised May 18, 2004.

L. T. Clark is with the Dept. of Electrical Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: lawrence.clark@asu.edu).

F. Ricci and M. Biyani are with Intel Corporation, Chandler, AZ 85226 USA (e-mail: franco.ricci@intel.com; manish.biyani@intel.com).

Digital Object Identifier 10.1109/JSSC.2004.840987

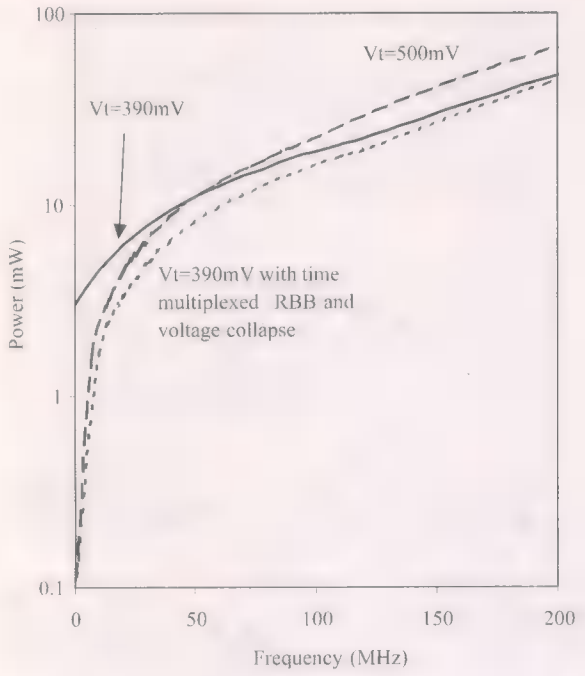


Fig. 1. Power utilizing RBB power-down modes interspersed with active operation versus no power-down mode and higher V_t.

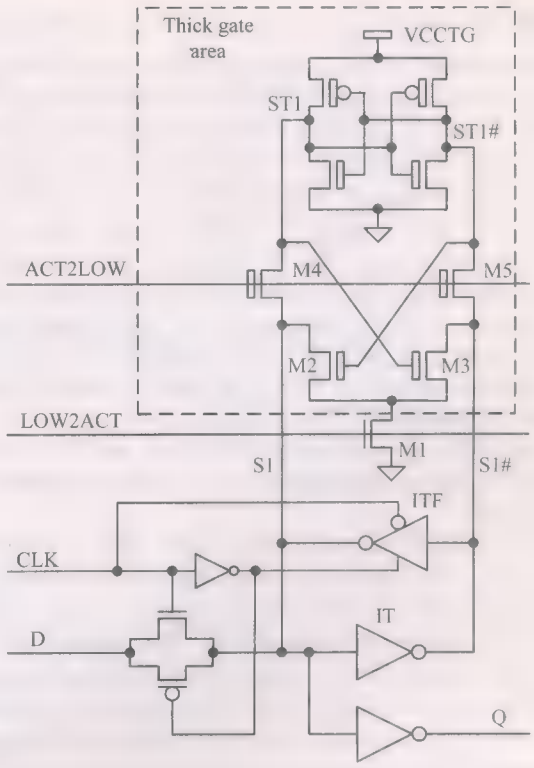


Fig. 2. Latch incorporating thick-gate state retention element. Thick-gate, high-V_t transistors are evident by the box gate symbols.

Section IV the addition of scan capability, and Section V comprises the experimental results and discussion. We conclude in Section VI. This paper focuses entirely on logic rather than memory usage, i.e., register file, latch, and flip-flop applications, while neglecting SRAM, although the use of thicker gate for SRAM has been explored [13].

II. Circuit Configuration and Process

The basic latch element is shown in Fig. 2 and comprises a thin-gate transistor high-performance latch comprised of the CMOS pass gate, feedforward inverter IT and feedback tri-state inverter ITF. The shadow thick-gate latch is comprised of transistors having both high V_t affording low I_{off}, and thicker oxide for low I_{gate}. The thick-gate region is outlined in the figure for clarity and the box gate symbols will be used to differentiate them from thin-gate transistors throughout this paper. This expands on the concept of high-V_t balloons described in [6] and the idea of maintaining supply power only to the state elements in an IC, while cutting off leakage to the combinational logic via MTCMOS schemes [7], [14]. The thick-gate portion is powered by a separate supply V_{CCTG}. The thick-gate transistors have higher V_t than the nominal thin-gate transistors, essentially severing the connection between low-voltage performance, maximum performance, and the standby power of the design. Early simulations showed that using the thick-gate transistors for the storage elements limited the register file write speed to less than 300 MHz if written in the phase before a read, while target designs included performance up to 2 GHz. Similarly, late data input to a transparent latch could result in an unacceptable timing push-out.

Our designs commonly use pulse-clocked latches to simulate master-slave flip-flops at lower power and size. Slower thick-gate transistors would require wider clock pulses and increase effort aimed at meeting hold requirements in timing convergence. This is described in detail in Section II-A. Consequently, the redundant latch scheme as shown in the figure was chosen, whereby the thick-gate write time, invoked only during entrance into a low-power state, does not limit operational speed. It is expected that the low-power state will be entered and exited at less than kHz rates, making the thick-gate write speed unimportant. For instance, in cell phone applications, the standby time can be on the order of seconds, between phone communications with the cell base stations.

In the processes used to validate the circuits described here, the thick-gate transistors support IO and analog circuitry, which traditionally use a higher V_t [15]. While a transistor optimized for this application would be preferable for electrical performance, it would increase process complexity and adversely affect die cost. Since storing the state in the separate latches requires no high voltages, the thick-gate transistors can be drawn at reduced channel length compared to their normal high-voltage design rules, to improve layout density. In practice, layout density is limited by the thick to thin-gate-oxide spacing.

During active operation V_{CCTG} is shorted to V_{CC} on die to limit IR drop induced noise between the supplies. Upon power-down, the state is first written to the thick-gate domain, then the entire combinatorial logic portion has the power supply removed as in MTCMOS. Rather than gate the V_{SS} supply node as done in our earlier RBB designs, the V_{CC} is removed externally at the regulator, mitigating the IR drop and die size associated with on-die power supply clamps.

A. Active Operation

As mentioned, to limit active power and delay through sequential elements, a pulsed-clock latch simulates a master-slave flip-flop (MSFF) as shown in the waveforms in Fig. 3. This has been shown to afford greater than 40% clock and sequential element energy savings as well as allowing some time borrowing to alleviate clock skew [11], [16]. The resulting sequential elements are smaller than a MSFF, helping to limit the overall sequential circuit size. These advantages are substantial enough to merit increased effort in designing to the greater hold times required. The signal LOW2ACT is de-asserted low in active mode operation, decoupling the thick-gate portion from the thin-gate high-performance latch. The minimal added capacitance due to the drains of thick-gate transistors M2–M5, which can be minimum sized, has a small effect on circuit speed and power in the active mode.

Fig. 3(a) shows the write timing of the pulse-clocked latch. The signal LOW2ACT is asserted low and so is not shown. The storage nodes S1 and S1# are quickly written, allowing a short clock pulse on signal PCLK. The timing used is for a 1.5 GHz design with the clock period shortened to provide margin for worst-case clock skew. Timing analysis is performed across process corners and voltages to determine the appropriate clock pulse (PCLK) width for each target process. Fig. 3(b) illustrates the slower response of using the thick gate alone. For these purposes, transistors M1-M3 and the feedback thin-gate tri-state inverter ITF in Fig. 2 are removed. Otherwise the circuit is unchanged. This makes the thick-gate latch, connecting nodes ST1 and ST1# the only state storage. Here, ACT2LOW is left enabled high so that nodes ST1 and ST1# can provide state storage in active operation. Note that thick-gate nodes ST1 and ST1# respond much more slowly and with the same pulse width the storage nodes fail to write even at 1.2 V V_{CC}. The design can only effectively pull down on the thick-gate storage node. This creates a slow transition, particularly when rising, since it is pulled up via the small thick-gate PMOS. The higher V_t of the thick-gate transistors will cause even further degradation in write timing at lower voltages. This makes use of thick-gate-only latches with DVS problematic.

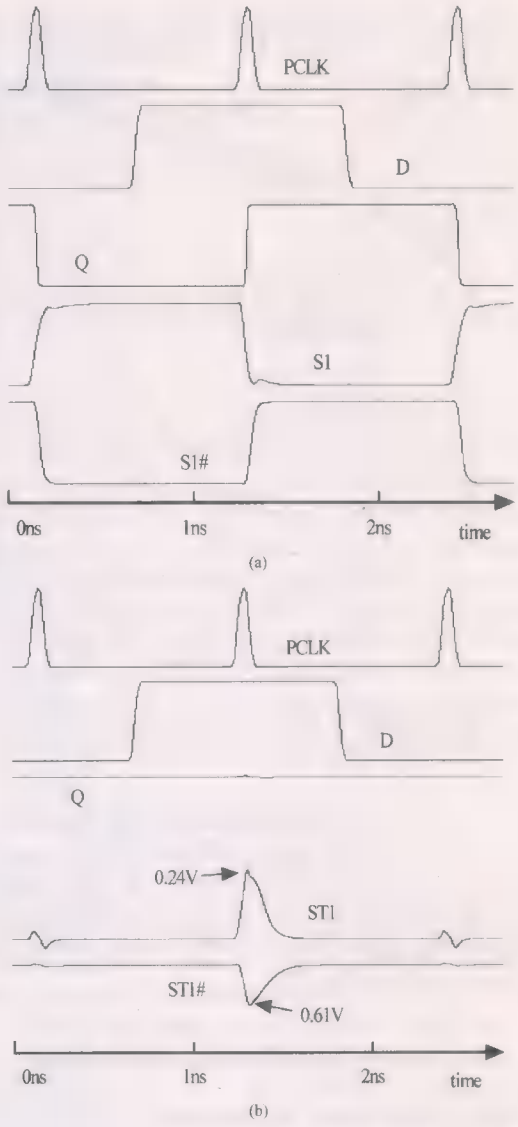


Fig. 3. Pulse-clocked latch timing (a) with LOW2ACT asserted low and (b) with LOW2ACT asserted high. All waveforms are 1.2 V amplitude except ST1 and ST1# that are marked.

B. Entering Standby Mode

To enter the low-standby-power mode, ACT2LOW is asserted high and the higher performance transistors in the thin-gate latch differentially write the thick-gate latch via the thick-gate pass transistors M4 and M5 as shown in Fig. 4. This operation relies upon the thin-gate devices having larger drive than the thick-gate devices. This is guaranteed by the lesser current drive of the thick-gate transistors due to their higher V_t, as well as by sizing. Of course, this must be simulated across process corners and the required voltage range at worst-case opposing data conditions, where both charge sharing and opposing currents may cause back writing. The thick-gate latches are all a single small size limited by the thick-gate design rules. Only the high-performance thin-gate transistors drive subsequent circuit stages.

Entrance into the low-standby-power state is completed by subsequent de-assertion of ACT2LOW to isolate the thick-gate latches. At this point, the V_{CC} can be floated or driven low by the external regulator. Floating the supply is preferred, since if the mode is exited soon after entrance, less power supply charge is needed to restore the operating supply voltage. The stored state is isolated via thick-gate transistors limiting the standby power to the leakage of the thick-gate storage elements. All N-wells are connected to V_{CCTG}, to avoid the increased size that well gaps would incur. The N-well leakage component is inconsequential. This also avoids discharging and charging the well capacitance when entering and leaving the low-standby-power mode.

The scheme disables all logic activity in the V_{CC} power domain and since the supply is floated, eventually leaking to 0 V, clocks are low while in this state. Thus, the entire clock tree can be on the main power supply and clock tree leakage is also eliminated. For a design predominantly using rising edge triggered flip-flops or pulse-clocked latches, it is then best to stop the clock in the low phase. Having resolved the majority of the

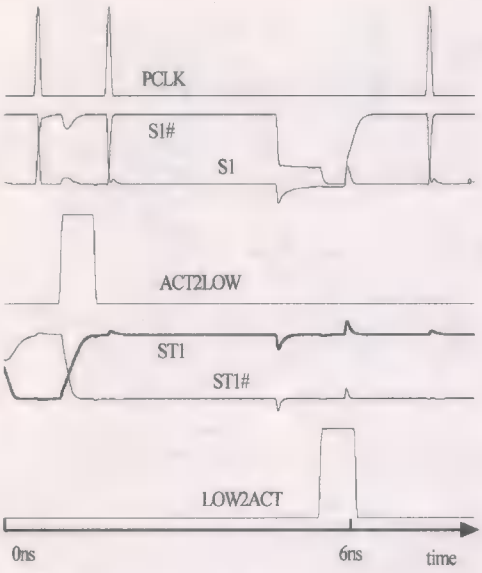


Fig. 4. Simulated operational waveforms, showing entrance into and exit from the low-power standby state.

cases with clock low, uncommon but very important cases are left and are discussed in Section III.

C. Exiting Standby Mode

To exit the low-standby-power mode, the signal LOW2ACT is asserted high, turning on M1 and providing a ground connection to transistors M2 and M3 that differentially sink current to set the state of the thin-gate latch upon power-up. As the supply increases from 0 V, the thick-gate transistors, having full gate overdrive of $V_{\mathrm{CCTG}} - V_t$, overpower the thin-gate transistors while they are in subthreshold operation. This forces the thin-gate storage to the correct state as it powers up, as in the ferroelectric shadow state storage for SRAMs described in [17], [18]. In the event that the supply does not completely collapse, the thin-gate latch state is not lost until the cell is sufficiently weak to allow writing via transistors M2 and M3. This case, where the V_{SS} does not fully collapse, is shown in Fig. 4, where the thick-gate transistors M2 and M3 must overpower the latch drivers I1 and IT1. Specifically, in Fig. 4 the thin-gate latch is purposely reversed after writing the state to the thick-gate shadow latch. The V_{CC} supply is then only collapsed to 300 mV. Nonetheless, the "one way" circuits correctly write the thin-gate latch state.

While it would have been possible to use the pass transistors M4 and M5 to write the thin-gate state during power-up [13], we found that this was less robust in the event of incomplete V_{CC} supply collapse combined with operation at process corners. Consequently, the "one-way" design shown was adopted despite the added size. To suppress I_{gate} due to transistors M2 and M3 while in standby, they must also be thick gate. Still more thick-gate transistors could have been added to make the write into the thick gate one-way as well, but due to the large drive difference to be expected between the thin and thick-gate transistors, easily ensured by proper sizing, this is unnecessary.

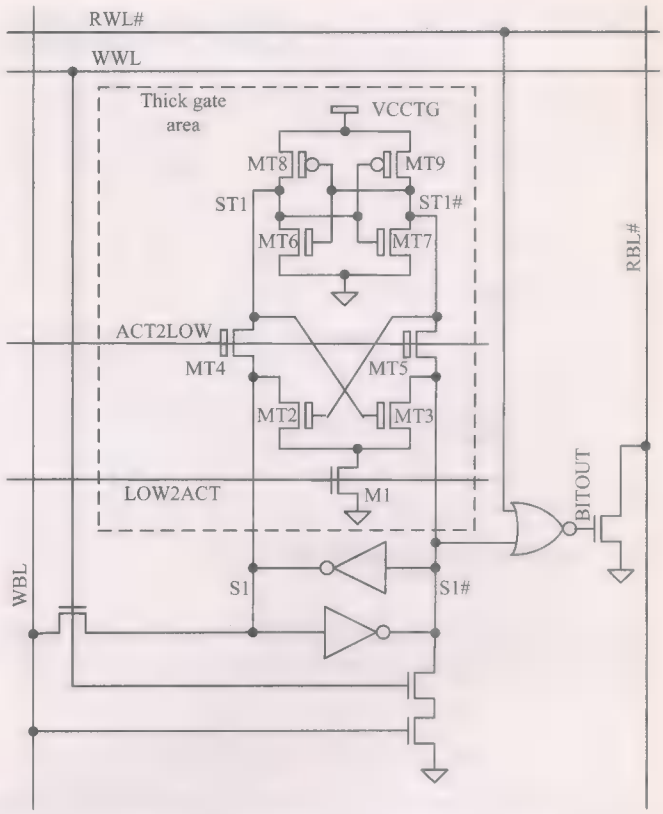


Fig. 5. Register file cell with thick-gate state retention devices.

D. Register Files

The register file design is shown in Fig. 5. The differential write assures good write performance at low voltages. The static NOR gate allows the pull down transistor to be half the width of a similar strength conventional stack. Thus, it lessens the dynamic domino read bitline (RBL#) load as well as limiting the leakage produced on this high fan-in dynamic node. It also increases the noise immunity to the read wordline (RWL# in the figure) by interjecting a static gate before the domino input transistor. The signal RWL# has less capacitive loading and overall read speed is retained.

As in the pulse-clocked latch, the thick-gate latch is not used as the primary storage node due to its slow speed. Specifically, when the register file is written late in the second phase of the clock and must be read in the next phase, a timing push out would occur if the nodes are incompletely written. The register file cell operation is illustrated in the simulation results comprising Fig. 6. The figure also includes three different write bitline (WBL) timings, separated by 50 ps. The thin-gate register file storage latch successfully writes with even with very late data setup time, analogous to the pulse-clocked latch case already described. Note that the latest WBL timing fails as shown by the failed write to node S1. By using only the thick-gate storage in the register file, ability to time-borrow, i.e., the late arrival of the write data in the write phase would have been sacrificed. Since the pertinent circuits are the same, operation when entering and leaving the standby mode is identical to the latch previously mentioned. Use of thick-gate-only storage

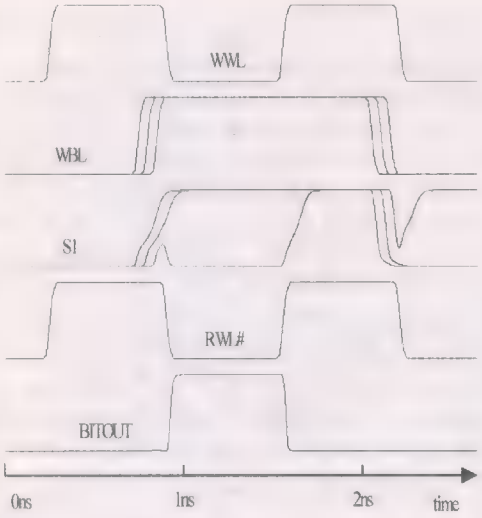


Fig. 6 Register file cell write with late data.

would have also limited operating speed to a clock phase length determined by the thick-gate latch write timing as mentioned.

III. Application to Other Circuits

The register file late write case is also applicable to the use of transparent high, rather than pulse-clocked latches. Since these allow time borrowing of nearly a clock phase, they are valuable for high-performance design. Shadow latches are attached just as in Fig. 2.

The use of thick-gate shadow latches is also applicable to master-slave flip-flops. Since the global clock is held low during standby as mentioned, the shadow latch is needed only on the slave latch. The master latch, in a transparent condition (due to clock low) during power-up will be set to the state required by preceding latches during state recovery. This limits the overhead significantly. It should also be noted that the slave has a half cycle to set the retaining element, since the write of that latch always occurs beginning at the clock rising edge. Alternatively, a flip-flop that is negative edge triggered requires that the shadow latch be attached to the master rather than the slave, so that the state properly propagates through the transparent on clock low latches as set by the shadow latch state.

A. Dynamic Logic

High-performance microprocessors frequently include a substantial amount of logic implemented by precharge-discharge dynamic (domino) logic. Even in lower performance designs, memories and register files are usually implemented in this style. Domino circuit paths must end in a dynamic to static conversion stage, typically a latch, which holds the output state through the domino circuit precharge phase. Therefore it is important to comprehend these circuits in any low-standby-power scheme.

A prototypical domino circuit is shown in Fig. 7. Here D2 (footless) domino stages with outputs A and B are combined in a NAND function by a set-dominant latch (SDL) that, besides the NAND, functions as the dynamic to static conversion latch

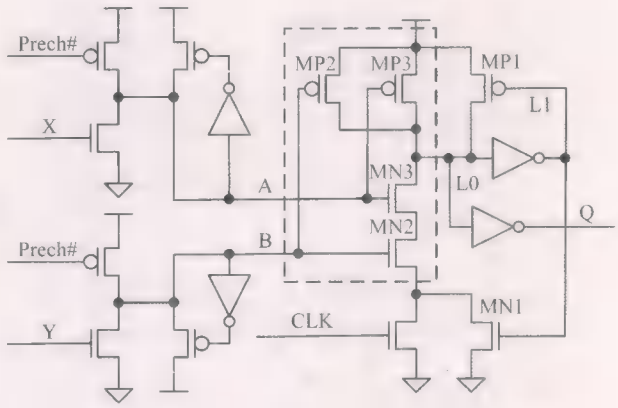


Fig. 7. Domino circuit and SDL dynamic to static conversion latch.

with minimal delay. A typical application would be that nodes A and B represent register file bitlines, with the read sense and latching function provided by the SDL. The NAND gate is outlined in the figure, where the additional transistors provide the latch and output driver function. Specifically, transistors MP1 and MN1 provide the latch function, with the latter creating a path to ground through transistors MN2 and MN3. Setting the latch node L0 to a one by asserting either nodes A or B low independent of the clock creates the set dominance.

As typically done to isolate the storage node L1 from the output, separate feedback and output inverters are used. This also allows different P to N ratios for the feedback and output inverters, separately optimizing read speed and noise immunity. In general, since the critical edge is L0 rising, the output inverter P to N ratio should be skewed to speed the falling edge at node Q. Noting that domino signals X and Y are only asserted high during clock high, nodes A and B can be asserted low during the same clock phase. Feedback transistor MN1 provides a path to ground while the clock CLK is low.

The timing is shown in Fig. 8. At the clock rising edge, the storage node L0 is discharged, since nodes A and B are precharged high in the previous (clock low) phase. When either node A or B is discharged low (only A is discharged in the figure), the latch immediately follows via the single PMOS pull-up transistor MP3 that comprise one of the two PMOS pull-ups MP2 and MP3 of the NAND gate. In Fig. 8, signal X rises in the clock high phase, discharging node A, which propagates to the output and is latched as shown. In keeping with the register file example, this corresponds to node BITOUT in Fig. 5, while node A corresponds to node RBL#, the read bitline in Fig. 5.

B. Dynamic Logic Standby Operation

Since the clock is held low in standby, all domino circuits that evaluate while the clock is high (phase 1 domino) are in the pre-charge state when entering and leaving the low-power mode and the set dominant latch (SDL) dynamic to static converter latch holds the previously evaluated state. By adding a thick-gate shadow latch to the SDL (see Fig. 9), the proper state is restored to the circuit before returning to active operation. Clock low (phase 2) domino circuits are evaluating upon

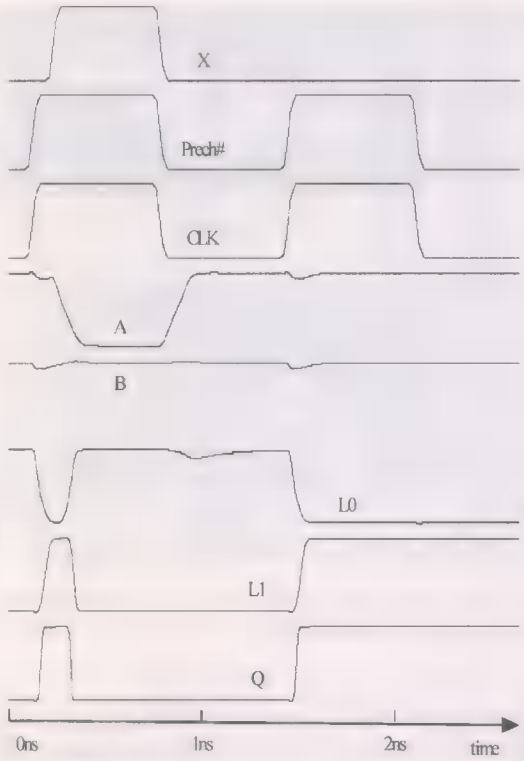


Fig. 8. Domino logic and SDL dynamic to static conversion operation.

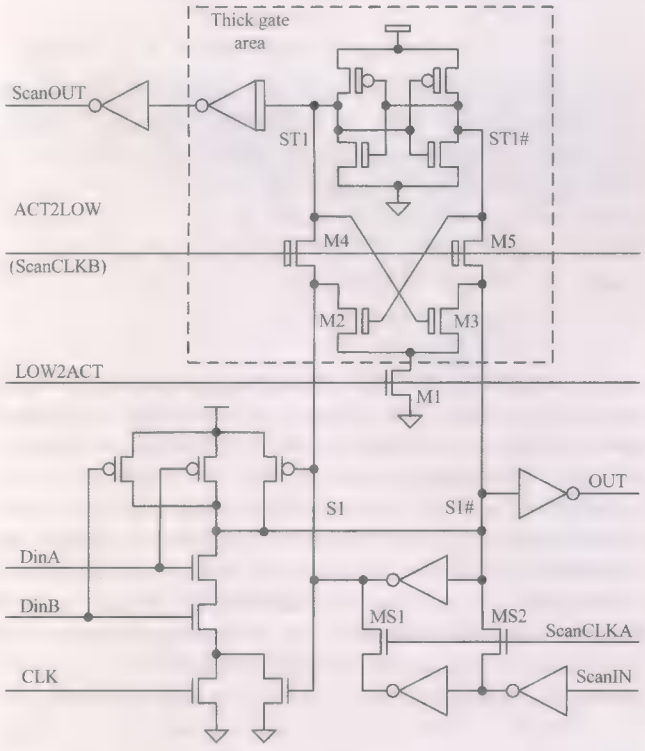


Fig. 9. Thick-gate NAND set-dominant latch dynamic to static converter with integrated thick-gate scan slave and state retention latch.

entrance to the low-power state. Hence, the half-latches comprised of their PMOS keepers may represent the proper state. This presents the problem of where to keep this state while in the low-power mode, as well as how to avoid falsely discharging

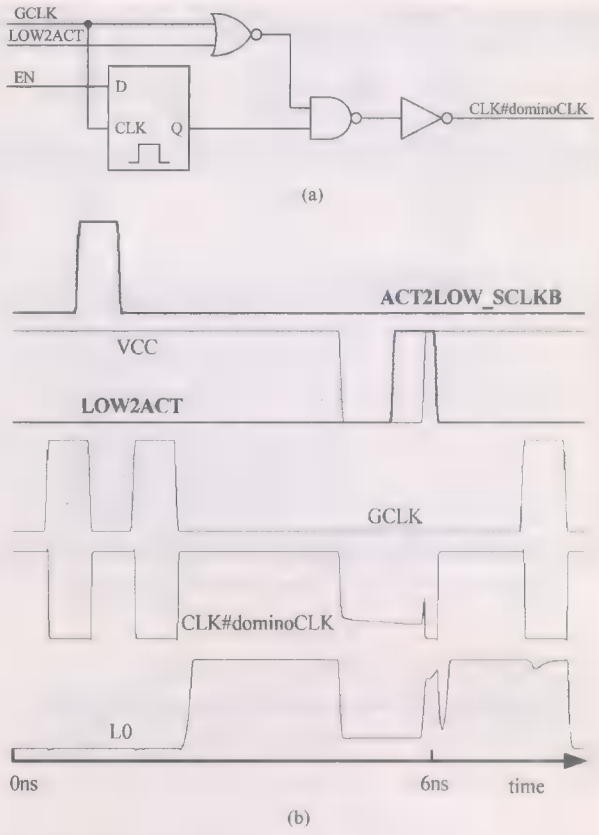


Fig. 10. Phase 2 (clock low evaluate) domino clock control (a) and simulation showing storage and reproduction of dynamic circuit state when entering and exiting the low-power state (b).

domino nodes that could in turn, disrupt downstream state nodes when the supply is restored.

C. Return From Standby for Dynamic Circuits

Precharging all evaluating domino while exiting the low-power state and subsequently allowing them to re-evaluate after return to the active state solves this problem. It also eliminates the possibility of erroneous domino operation at very low V_{CC}, where the sum of NMOS transistor off currents may become comparable with the keeper on current. This condition will cause the local domino node half-latch to be upset. The precharge and re-evaluate is accomplished by using the return to active signal, LOW2ACT, to enable the local clock buffer used for clock low domino as shown in Fig. 10(a). The low-phase domino clock, CLK#dominoCLK is forced low while LOW2ACT is high, forcing the domino node into precharge as power is restored, as evident in the figure. When LOW2ACT falls, this clock rises causing the domino gates to evaluate before active operation begins. This clock assertion is simply forced by the LOW2ACT signal input to the NOR gate. Thus, the domino inputs are set by the shadow latches, and the domino nodes are returned to their proper state by the single evaluate clock edge, independent of the state that they powered up in or collapsed to under a low supply voltage condition.

Fig. 10(b) shows a circuit simulation of this operation. The signal ACT2LOW_SCLKB is asserted to write the thick-gate shadow latch as before. Another clock cycle alters the state of the domino node and the supply is subsequently collapsed,

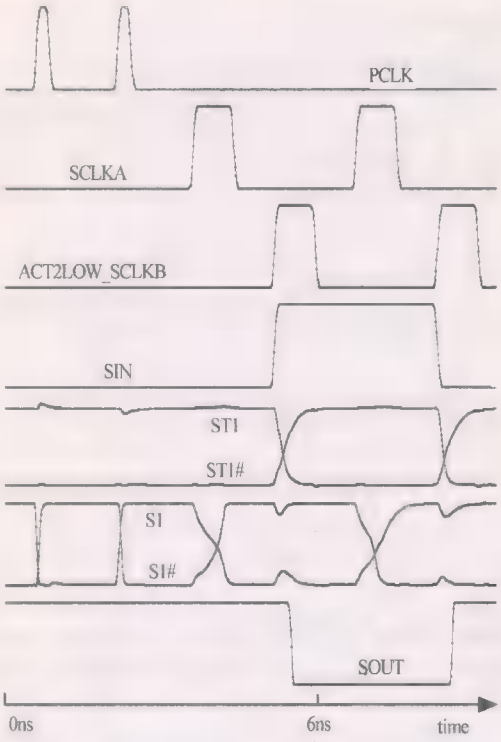


Fig. 11. Scan mode circuit operation.

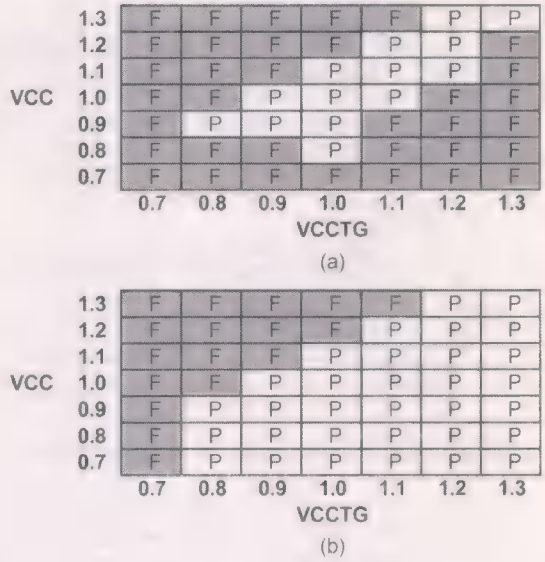


Fig. 12. Shmoo plots showing operational voltages. In (a) the voltages are imbalanced while entering and exiting the low-power mode and in (b) only upon exit.

also as before. LOW2ACT is asserted while the thin-gate power supply V_{CC} is returned high. This precharges the domino gate by forcing the active high clock signal CLK#dominoCLK low. The rising edge of CLK#dominoCLK re-evaluates the domino gate with inputs driven by the preceding shadow state. The SDL latch node L0 is shown to follow the evaluate, including the glitch due to precharge propagation. The clock then resumes with the clock low evaluate domino gates in the correct state.

IV. Scan Design

By requiring an extra latch the scheme increases the overall circuit area as mentioned. However, using the shadow latch as the scan slave as illustrated in Fig. 9 can mitigate the area increase. To limit the increase in loading on the high-performance latch, it is written differentially in scan. Separate scan clocks allow nonoverlapping clock operation in scan, using the SCANCLKA and SCANCLKB (ACT2LOW) signals. This allows looser routing of the scan clock signals, which can be treated by routers as signals rather than clocks, as well as eliminating race-through conditions on the scan chain. No separate scan enable signals are required. Operation is shown in Fig. 11. Referring to Fig. 9, signals DinA and DinB are high (held in precharge) and CLK is held low. The data is then scanned into the thin-gate latch by asserting SCANCLKA, and into the thick-gate slave by asserting ACT2LOW_SCANCLKB, respectively as shown. Race through risk during scan is also lessened by the relatively low performance of the thick-gate slave latches, but limits scan operation to 300 MHz for the reasons described previously (note the slow latch transitions). This limitation should not have significant effect on test time or usability of the scan feature. Since there are few extra signals and supplies and given that auto-placed and routed logic blocks are generally wire limited, the impact on block size is minimal in that case. For register files, which do not require scan capability, some of the size impact can be limited by placing the thick-gate devices under metal limited portions of the cell and is inversely proportional to the number of ports.

V. Experimental Results

A test die containing a 32 entry translation lookaside buffer built from the register file cells (as well as CAM cells) and four scan chains of 2000 pulse-clocked latches each, containing at total of 9920 thick-gate state retention latches, was fabricated in a 90-nm process. A die plot is shown in Fig. 13, where the test structure is $600 \times 1700\ \mu\mathrm{m}$ and the active circuits are $0.51\ \mathrm{mm}^2$ in area. Fig. 12(a) is a Shmoo plot showing the passing and failing voltages on the thin and thick-gate domains. At the intended operating point, i.e., $V_{CC} = V_{\mathrm{CCTG}}$, successful operation is shown down to 0.8 V. As the voltage on the thick-gate domain is raised above the thin-gate domain, charge sharing can cause the write to fail, upsetting the thin-gate domain rather than writing the thick-gate domain. At high V_{CC} and low V_{CCTG}, the thin-gate transistors and large capacitance of the thin-gate latch overpower transistors M1-M3, so the correct state cannot be written back. In actual usage V_{CC} will be strictly equal to or lower than V_{CCTG} due to leakage. In the former case, the state is retained and in the latter case, correct operation has been confirmed as in the simulation results described in Fig. 4 and shown in Fig. 12(b). The test die is comprised of minimum sized latches, while a real design will use a mix of large and small latches. Larger latches are less susceptible to back writing, so the measured results constitute a worst-case.

The thin-gate threshold voltages were measured on e-test structures to be $V_{\mathrm{tn}} = 413\ \mathrm{mV}$ and $V_{\mathrm{tp}} = -456\ \mathrm{mV}$, while the thick-gate threshold voltages were $V_{\mathrm{tn}} = 900\ \mathrm{mV}$ and $V_{\mathrm{tp}} = -760\ \mathrm{mV}$. Both sets of values are higher than the targets. The very large thick-gate values and imbalance in particular, account for the relatively high measured minimum

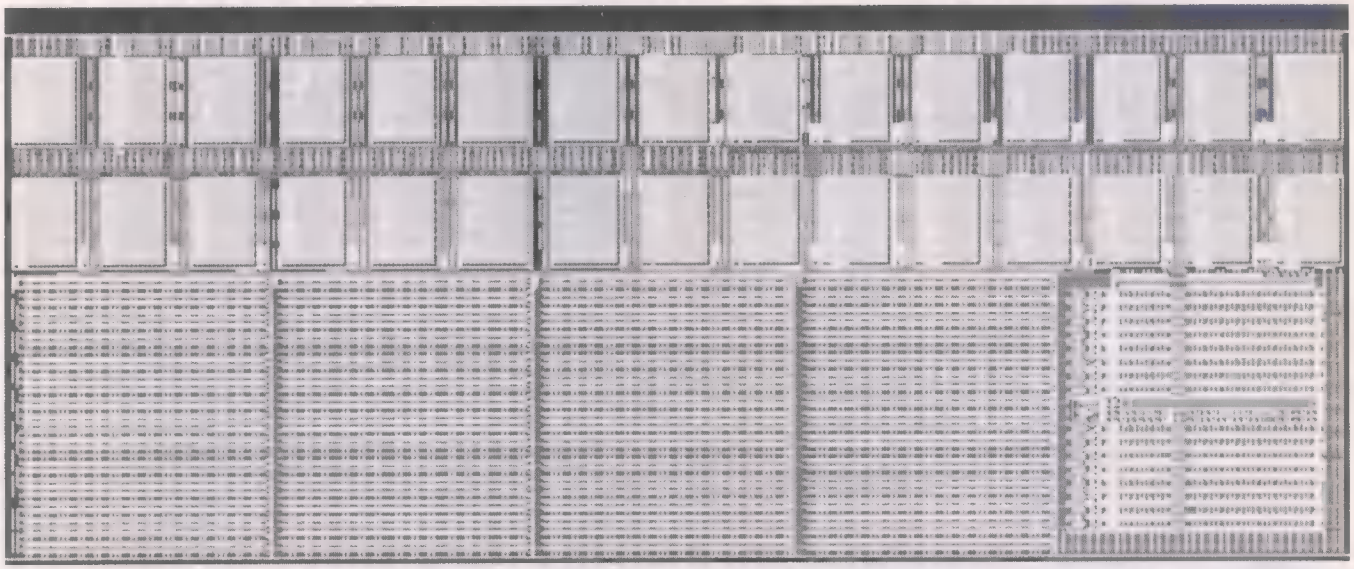


Fig. 13. Die plot of the test chip. The four shift register arrays as well as the TLB are evident left to right.

operating voltage (V_{CCMIN}), by causing the write to be subthreshold through the thick-gate NMOS series transistors. The thick-gate threshold voltages can be lowered substantially without affecting the standby power, allowing improved V_{CCMIN}. Additionally, lower thin-gate threshold voltages will improve active power while not affecting that in standby.

Measured total power supply current on V_{CCTG} in the low-power state was between 2 and 6 mA, corresponding to 202 to 606 pA per cell, respectively, depending on the die and voltage, at room temperature. This is attributable to aggressive halo doping and consequent band-to-band tunneling at the drain edges. Depending upon process architecture, this can be lowered substantially. It may also be addressed via design by lowering the V_{CCTG} supply voltage while in the low-power state. This will be a topic of future work.

VI. Discussion and Conclusions

Standby leakage presents a considerable obstacle to transistor scaling for future battery operated devices. We have presented a latch design that allows low standby power for sub 130-nm processes, which have gate leakages that in and of themselves exceed typical 100 μA standby limits for an IC. The number of transistors in the design is limited, helped in large part by the use of pulse-clocked latches rather than master-slave flip-flops. This choice improves performance, energy and size. For instance, the master-slave design of [6] requires 32 transistors while this design requires only 21. The previous design is also prone to charge sharing during power-up, while in the design presented here, a one-way write to the thin-gate high-performance domain alleviates any possibility of back-writing in the event of incomplete supply collapse. In our design, the latch speed is optimized for high performance by bypassing the thick-gate high-V_t transistors during active operation while low standby power is achieved by storing the state in low-leakage transistors. The approach has been shown to be applicable to a wide range of static and dynamic circuits. Finally, the added size due to the larger thick-gate transistors and increased spacing between thin and thick-gate devices is effectively mitigated by using the thick-gate storage element as the scan slave. Non-overlapping clocks in the scan mode of operation alleviates race-through.

Acknowledgment

The authors thank the technology development groups for their contributions, as well as J. Heeb for management support.

References

[1] S. Thompson, I. Young, J. Greason, and M. Bohr, "Dual threshold voltages and substrate bias: Keys to high performance, low power, 0.1 μm logic designs," in *VLSI Symp. Tech. Dig.*, 1997, pp. 69–70.

[2] H. Mizuno *et al.*, "An 18-μA standby current 1.8-V, 200-MHz microprocessor with self-substrate-biased data-retention mode," *IEEE J. Solid-State Circuits*, vol. 34, no. 11, pp. 1492–1500, Nov. 1999.

[3] L. Clark, N. Deutscher, F. Ricci, and S. Demmons, "Standby power management for a 0.18-μm microprocessor," in *Proc. Int. Symp. Low Power Electronics and Design*, 2002, pp. 7–12.

[4] K. Osada, Y. Saitoh, E. Ibe, and K. Ishibashi, "16.7 fA/cell tunnel-leakage-suppressed 16 Mb SRAM for handling cosmic-ray-induced multi-errors," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2003, p. 302.

[5] S. Mutoh *et al.*, "1 V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.

[6] S. Shigematsu, S. Mutoh, Y. Matsuya, Y. Tanabe, and J. Yamada, "A 1-V high-speed MTCMOS circuit scheme for power-down application circuits," *IEEE J. Solid-State Circuits*, vol. 32, no. 6, pp. 861–870, Jun. 1997.

[7] V. Zyuban and S. Kosonocky, "Low power integrated scan-retention mechanism," in *Proc. Int. Symp. Low Power Electronics and Design*, 2002, pp. 98–102.

[8] D. Frank, "Power constrained CMOS scaling limits," *IBM J. Res. Develop.*, vol. 46, no. 2/3, pp. 235–235, 2002.

[9] D. Frank *et al.*, "Device scaling limits of Si MOSFET's and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, Mar. 2001.

[10] S. Thompson *et al.*, "MOS scaling: Transistor challenges for the 21st century," *Intel Technology J.*, vol. Q3, pp. 1–15, 1998.

[11] L. Clark *et al.*, "An embedded 32-b microprocessor core for low-power and high-performance applications," *IEEE J. Solid-State Circuits*, vol. 36, no. 11, pp. 1599–1608, Nov. 2001.

[12] A. Keshavarzi *et al.*, "Technology scaling behavior of optimum reverse body bias for leakage power reduction in CMOS IC's," in *Proc. Int. Symp. Low Power Electronics and Design*, 1999, pp. 252–254.

[13] L. T. Clark and F. Ricci, "Low standby power using shadow storage," U.S. Patent 6,639,827, Oct. 28, 2003.
[14] L. Clark, "Trends and challenges for wireless embedded DSP's," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2003, pp. 171–176.
[15] K. Kuhn *et al.*, "A 90 nm communication technology featuring SiGe HBT transistors, RF CMOS, precision R-L-C RF elements and 1 μm^2 6-T SRAM cell," in *IEDM Tech. Dig.*, 2002, pp. 73–76.
[16] J. Tschanz *et al.*, "Comparative delay and energy of single edge-triggered and dual edge-triggered pulsed flip-flops for high-performance microprocessors," in *Proc. Int. Symp. Low Power Electronics and Design*, 2001, pp. 147–152.
[17] S. Sheffield, D. Eaton, M. Butler, D. Parris, H. Wilson, and A. McNeillie, "A Ferroelectric Nonvolatile Memory," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Mar. 1988, pp. 130–131.
[18] T. Miwa *et al.*, "NV-SRAM: A nonvolatile SRAM with backup ferroelectric capacitors," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 522–527, Mar. 2001.
[19] L. Clark, M. Morrow, and W. Brown, "Reverse body bias and supply collapse for low effective standby power," *IEEE Trans. Very Large Scale Integrated (VLSI) Syst.*, vol. 12, no. 9, pp. 947–956, Sep. 2004.

Lawrence T. Clark (M'90–SM'2001) was born in Detroit, MI. He received the B.S. degree in computer science from Northern Arizona University, Flagstaff, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, in 1987 and 1992, respectively.

He worked at Intel Corporation, in 1982 and from 1984 to 1985 in product and test engineering, and at VLSI Inc. from 1990 to 1992 in chipset design. From 1992 to 2003, he worked at Intel Corporation, in various capacities including microprocessor design, (participating in Pentium, Itanium, and XScale processor designs), compact modeling for circuit simulation, and CMOS imager design. Most recently, he was a Principal Engineer and Circuit Design Manager for XScale Microprocessors. In 2003, he joined the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, as an Associate Professor. He holds over 40 patents and has approximately 20 pending. His research interests are circuits, architectures, and computer-aided design for high-performance and low-power VLSI systems. In August 2004, Prof. Clark joined the Department of Electrical Engineering at Arizona State University, Tempe, AZ.

Franco Ricci was born in Detroit, MI. He received the B.S. degree in electrical engineering from the University of Michigan, Ann Arbor, in 1991 and the M.S. degree in electrical engineering from Arizona State University, Tempe, in 1997.

Since January 1992, he has been with Intel Corporation, Chandler, AZ, in various capacities including microprocessor design (participating in Itanium and XScale processor designs), and design automation. His current interests are in the fields of low-power microprocessor design and performance verification.

Manish Biyani received the B.S. degree in electronics and communications engineering from R.E.C. Trichy, India, and the M.S. degree in electrical and computer engineering from the University of Florida, Gainesville.

He has worked at Intel Corporation in IC design for the last seven years. He has worked on several XScale microprocessor cores during this period. His areas of expertise include high-speed/low-power digital datapath and memory circuit design, as well as developing design methodologies. He currently holds one patent.

A High-Performance Very Low-Voltage Current Sense Amplifier for Nonvolatile Memories

Antonino Conte, Gianbattista Lo Giudice, Gaetano Palumbo, *Senior Member, IEEE*, and Alfredo Signorello

***Abstract*—A high-performance sense amplifier for nonvolatile memories capable of working under a very low-voltage power supply is presented. The topology of the sense amplifier uses a pure current-mode comparison allowing power supplies lower than 1 V to be used and includes two subcircuits which improve slew rate performance.**

The sense amplifier was implemented in an EEPROM realized with a 0.18-μm EEPROM technology. Experimental results showed a read access time of about 30 ns with a power supply of 1.65 V.

***Index Terms*—Current mode, EEPROM, low voltage, nonvolatile memory, sense amplifier, smart card.**

I. Introduction

VARIOUS electronic systems used in telecommunications (pagers, mobile telephones, etc.), in consumer products (smart cards, palmtops, digital video cameras and cameras), and in personal computers (BIOS) require nonvolatile memories with high speeds in both read and write operation modes as well as low power consumption [1]–[3]. The need for very low power consumption, which increases battery life time and portability, has become a key design aspect particularly for portable electronic equipment. To satisfy the low power constraints in the digital circuit domain, the customary way is to reduce the power supply voltage [4]–[12]. Hence, a 1.5-V-only (or even lower) nonvolatile memory is required in keeping with present voltage reduction trends [13]–[17].

An important example of portable microelectronics systems are Smart Cards, which have become of daily use in the last few years. Smart Cards, usually of the same dimension as credit cards and made of plastic materials, incorporate a microsystem containing several electronic subsystems that allow elaboration and memorization operations [18]. Contactless Smart Cards that derive their power supply from radio signals have become a trend [19], [20]. In this type of application, low-voltage nonvolatile memories, and in particular EEPROM, are needed. Moreover, given that the time interval when the Card is supplied is quite limited, the memories adopted must have extremely high read and write ratings. These requirements are difficult to satisfy when the objective is also to lower the supply voltage [13].

Manuscript received January 30, 2004; revised September 14, 2004.

A. Conte, G. Lo Giudice, and A. Signorello are with ST-Microelectronics, MPG Group, I-95121 Catania, Italy.

G. Palumbo is with the Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi (DIEES), Universita'di Catania, I-95125 Catania, Italy (e-mail: gpalumbo@diees.unict.it).

Digital Object Identifier 10.1109/JSSC.2004.840985

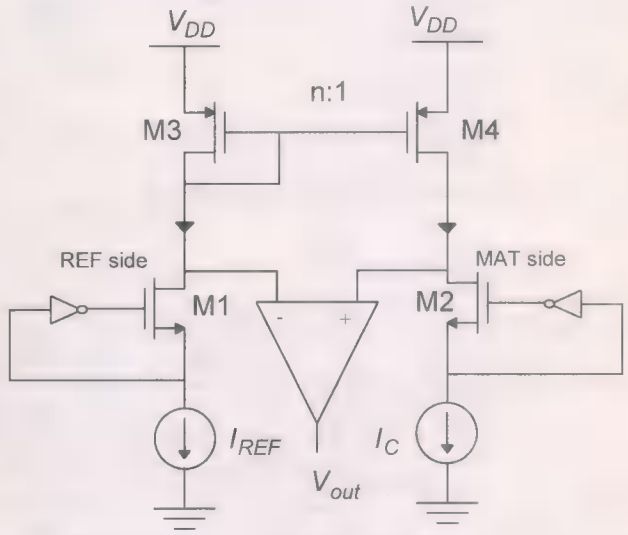


Fig. 1. Block scheme of a conventional sense amplifier.

Read speed is mainly determined by the read path, which is affected in a nonnegligible way by the sense amplifier's speed performance, and becomes critical when the power supply is reduced [21].

This paper focuses on a novel topology sense amplifier for nonvolatile memories, capable of operating at voltages as low as 1 V, and satisfying speed constraints. This sense amplifier operates under very low voltage without needing special low threshold voltage devices. The pre-charging speed performance of the bitline is still preserved, despite avoiding recourse to cascoding techniques for the pre-charge scheme to overcome low power supply limitations. These two features make the proposed scheme particularly appealing in standard memory processes and very low-voltage range of applications.

II. Sense Amplifier for Nonvolatile Memories

The reading operation of an EEPROM or Flash is performed by sensing the current cell under well-defined biasing conditions. In particular, a programmed EEPROM cell has a low threshold voltage, giving a high level current under the bias condition. In contrast, an erased EEPROM cell has a high threshold voltage, giving a low level current. The convention for a Flash memory cell is reversed. Read operation can clearly be achieved by comparing the current cell with a reference current generally provided by another cell normally linked with the process characteristics. Although the natural read operation can be performed in a current mode approach, traditionally a voltage mode operation is adopted. In fact, read operation is implemented by

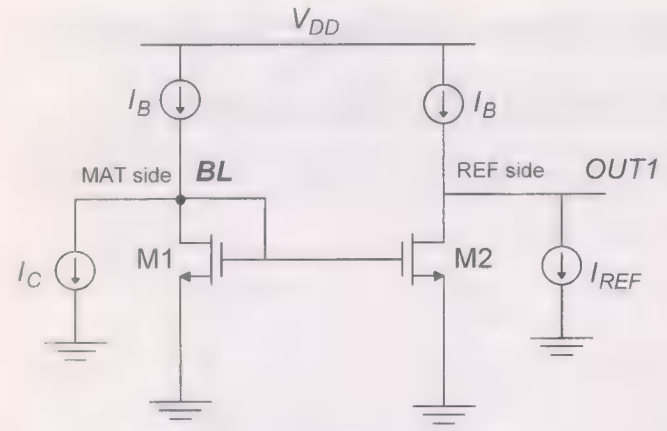


Fig. 2. Block scheme of a pure current mode comparison.

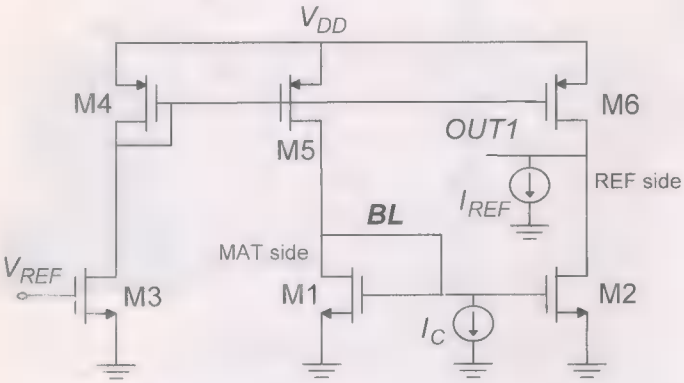


Fig. 3. Adopted implementation of the block scheme in Fig. 2.

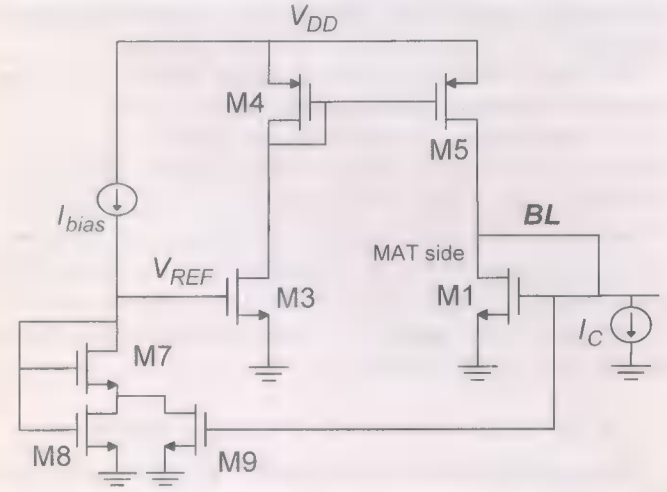


Fig. 4. Circuit topology to improve the slew rate of the sense amplifier.

using a voltage sense amplifier, which compares the voltage after the current is converted to voltage (Fig. 1).

In general, differential sense topologies, which have greater advantages than the corresponding single-ended version, are used. Their classic topology is based on the conventional block scheme in Fig. 1, where I_C and $I_{\rm REF}$ model the cell current and the reference current, respectively, and $V_{\rm OUT}$ is the sense amplifier output voltage [2], [21]. The current mirror M3–M4, with a mirror aspect ratio lower than one (and typically set to 0.5 to ensure an equal delay for a 1 or 0 read), is used to

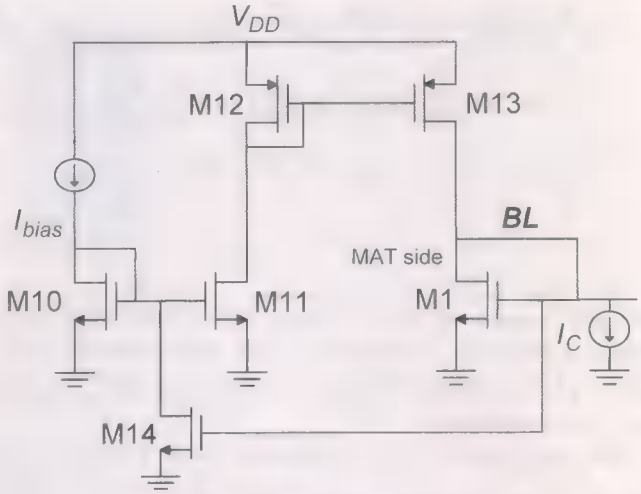


Fig. 5. Circuit block to improve the pre-charge phase of the sense amplifier.

TABLE I
TRANSISTOR ASPECT RATIOS

Transistor	W/L (μm)
M_1= M_2= M_3	2.2/1.1
M_4 =M_5= M_6=M_9	4.4/0.88
M_7	2.2/0.88
M_8	2.2/3.08
M_{10}=M_{14}	3.3/0.88
M_{12}=M_{13}	3.08/0.88
M_{15}= M_{16}=M_{11}	2.2/1.1
M_{17}= M_{18}	4.96/0.88
M_{19}= M_{20}	1.36/0.2

appropriately scale the current of the reference cell. Since a fundamental role is played by the pre-charging method in any sensing scheme for non volatile memories, traditionally this task has been accomplished by adopting cascoding techniques using an inverter with a source follower output stage and a unitary feedback loop. This approach allows fast pre-charging independently of the capacitive load represented by the bitline of the array (MAT side in the sensing scheme). Although this solution is very useful down to a power supply of 1.8-V, it shows nonnegligible limitations once the power supply is lowered further. This occurs because the source follower does not correctly bias the bitline at the desired level (imposed by technology constraints) and also affects the reading speed performance.

Although the block scheme in Fig. 1 is not suitable for low-voltage operation, introducing some of the modifications proposed in literature can allow its use with low-voltage nonvolatile memories, albeit at the cost of reducing performance [14]–[16].

In particular, the solution proposed in [14] based on the so-called self-biasing bitline sensing scheme, exploits the charge sharing effect between the dummy bitline (one for every sense amplifier) and the addressed bitline. In addition, it uses an n-channel transistor in cascoding configuration to separate the capacitive net of the bitline from the net used to perform the comparison. There are two drawbacks to this solution. One is control of the final bitline pre-charge level, which depends on the power supply (usually half the power supply), which is

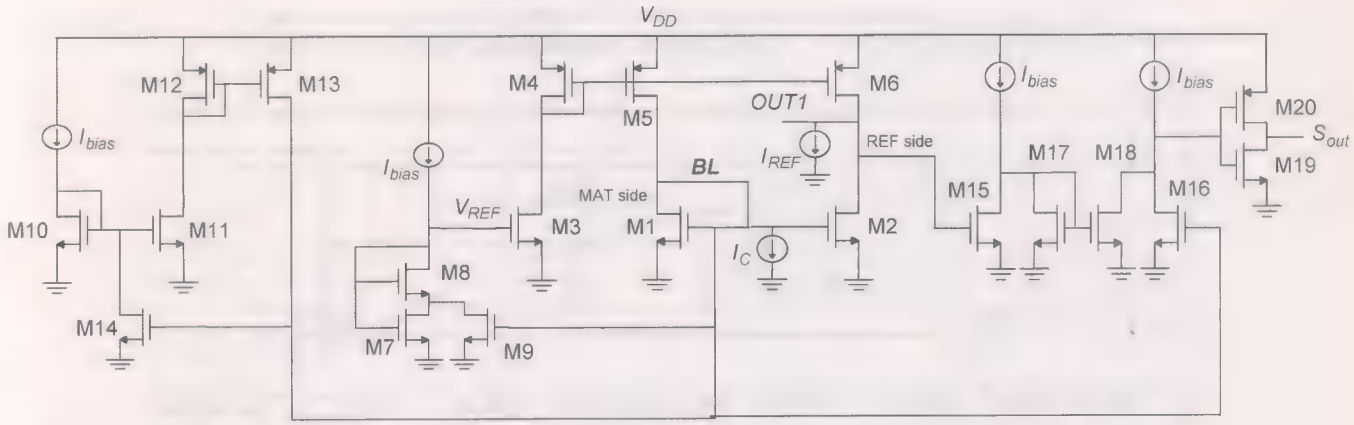


Fig. 6. Detailed scheme of the low-voltage sense amplifier.

TABLE II
SIMULATION RESULTS

V_{DD}	Temperature	Read access time	Minimum current compared
1.65 V	27° C	25 ns	0.5 μA
1.35 V	125° C	43 ns	1 μA
1.95 V	-40° C	15 ns	1 μA
1 V	27° C	38 ns	0.5 μA

itself not under control. The other is the need for an n-channel transistor to implement the cascoding of the bitline, thereby restricting very low-voltage operation.

The solution presented in [15] requires special low threshold voltage transistors which are not strictly mandatory, even if they can profitably used in other memory subcircuits (such as charge pumps). Another drawback is represented by the control of the bitline voltage biasing, which is not suitable for power supply voltages much lower than 1.5 V.

III. VERY LOW-VOLTAGE SENSE AMPLIFIER

The key idea behind the proposed topology is based on implementing a true current comparison operation [22], [23]. Current comparison is performed simply by a current mirror loaded with a current generator. According to the block scheme in Fig. 2, where OUT1 is sense amplifier output voltage, I_C and I_{REF} are the cell current and the reference current, respectively, and I_B is a bias current, the output voltage tends to the power supply when I_C, is greater than I_{REF}, otherwise it tends to ground. In particular, for small differences between I_C and I_{REF}, the output voltage swing around the bias condition, is given by

$$\Delta V_{\mathrm{out}} = r_{\mathrm{out}}(I_C - I_{\mathrm{REF}}) \tag{1}$$

where r_{out} is the small-signal resistance at the output node. Of course, the mirroring behavior disappears when current differences produce huge voltage swings. The output voltage becomes equal to the power supply or ground, as transistor M2 is forced to work in cut off or in the linear region, respectively.

A. Sense Amplifier Core

The drawbacks of the simple block scheme in Fig. 2 are due to the bias voltage required on the bitline node (i.e., node BL in Fig. 2) before the memory cell is connected (i.e., before the current I_C is applied). This must be accurately set to around 0.8 V, because it coincides with the drain node of the EEPROM cells and hence affects the cell current being sensed.[1] In particular, with typical current bias values, threshold voltage and process parameters, a minimum transistor size cannot be used. To overcome this drawback, and define the bias voltage on the bitline in a sufficiently insensitive manner, the block scheme in Fig. 3 was adopted. It is based on a p-type current mirror which sets on the diode connected NMOS transistor the same current which flows in an equally sized NMOS transistor with the required bias voltages on its gate. As shown in the Appendix, after setting transistors M3 and M5 equal to M1 and M4, respectively, and neglecting channel length modulation, as well as short channel effects, the voltage on the bitline when the memory cell is not connected (i.e., with $I_C = 0$) equals the reference voltage, V_{REF}.

The circuit in Fig. 3 maintains the low voltage features of the current mirror scheme in Fig. 2. Indeed, it can work with a power supply as low as a threshold voltage plus a saturation drain-source voltage, which with modern technologies means a value lower than 1 V. Under this extremely low-voltage power supply the drawback is the substantial difference between the drain-source voltages of the two transistor couples M1, M3, and M4, M5, that determines a non negligible error between the voltage reference and the resulting bitline voltage. However, as can be simply derived from the relationships included in the Appendix, with power supply voltages around twice the minimum power supply (i.e., $2V_T + 2V_{\mathrm{DS\ sat}}$, where $V_{\mathrm{DS\ sat}}$ is the drain-source saturation voltage of a transistor) an ideal matching condition

[1]Remember that the level of current of a nonvolatile cell under the different conditions (erased and programmed) changes varying the voltage drain, and technology is generally characterized for only a typical value.

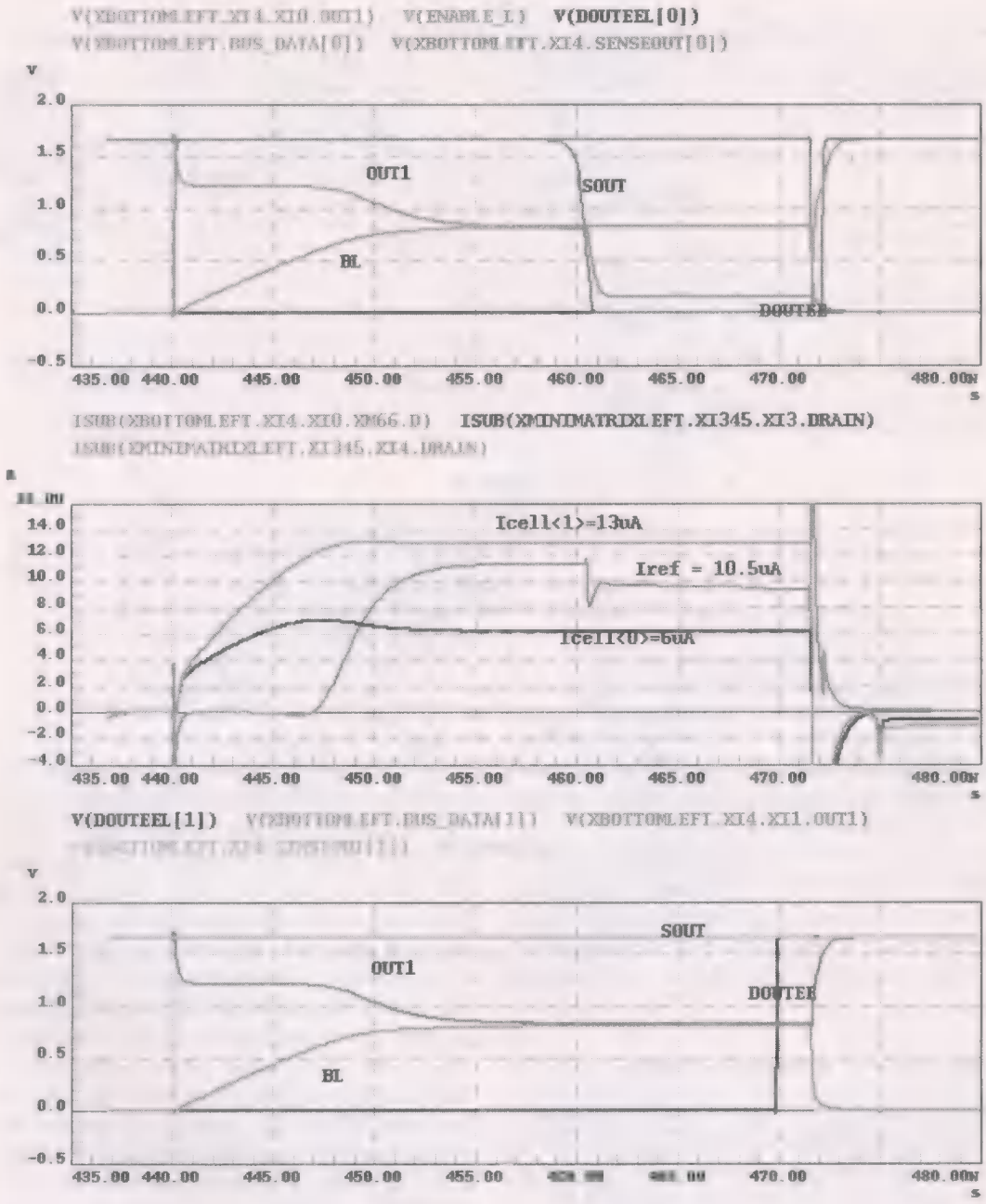


Fig. 7(a). Simulation results of the sense amplifier under a 1.65-V power supply assuming the cell deleted with a cell current equal to 6 μA (upper plot), the cell programmed with a current cell equal to 13 μA (lower plot).

between the drain-source voltages can be achieved by cancelling the channel length modulation effects. Thus, the resulting bitline voltage is ideally equal to the reference voltage.

B. Slew Rate Increase

Although the scheme in Fig. 3 is simple and efficient it exhibits the typical drawback of a limited slew rate (afflicting any class A amplifier) which limits speed performance to pre-charge the bitline at the required voltage. Indeed, the bitline represents a heavy capacitance load, C_{BL}, and the time slot required to charge it at the bias voltage, V_{REF}, is equal to

$$t_{\mathrm{pre}} = \frac{C_{\mathrm{BL}}}{I_3} V_{\mathrm{REF}} \tag{2}$$

where I_3 is the saturation current of M3. It is evident that to reduce the pre-charge time we need to increase the bias current. However, this proportionally increases power consumption.

To overcome this shortcoming, two adjoining circuits were added which increase the current so it charges the bitline only in the pre-charge time slot. The former, reported in Fig. 4, increases the reference voltage V_{REF} at the gate of transistor M3, which progressively decreases until the required value is reached. In particular, when the bitline is discharged, transistor M9 is switched off and transistor M7 and M8, which have equal width, become equivalent to a diode connected transistor with the same width and a length equal to the sum of the lengths of M7 and M8. Note that under this condition M7 and M8 work in saturation and the triode region, respectively. During the initial

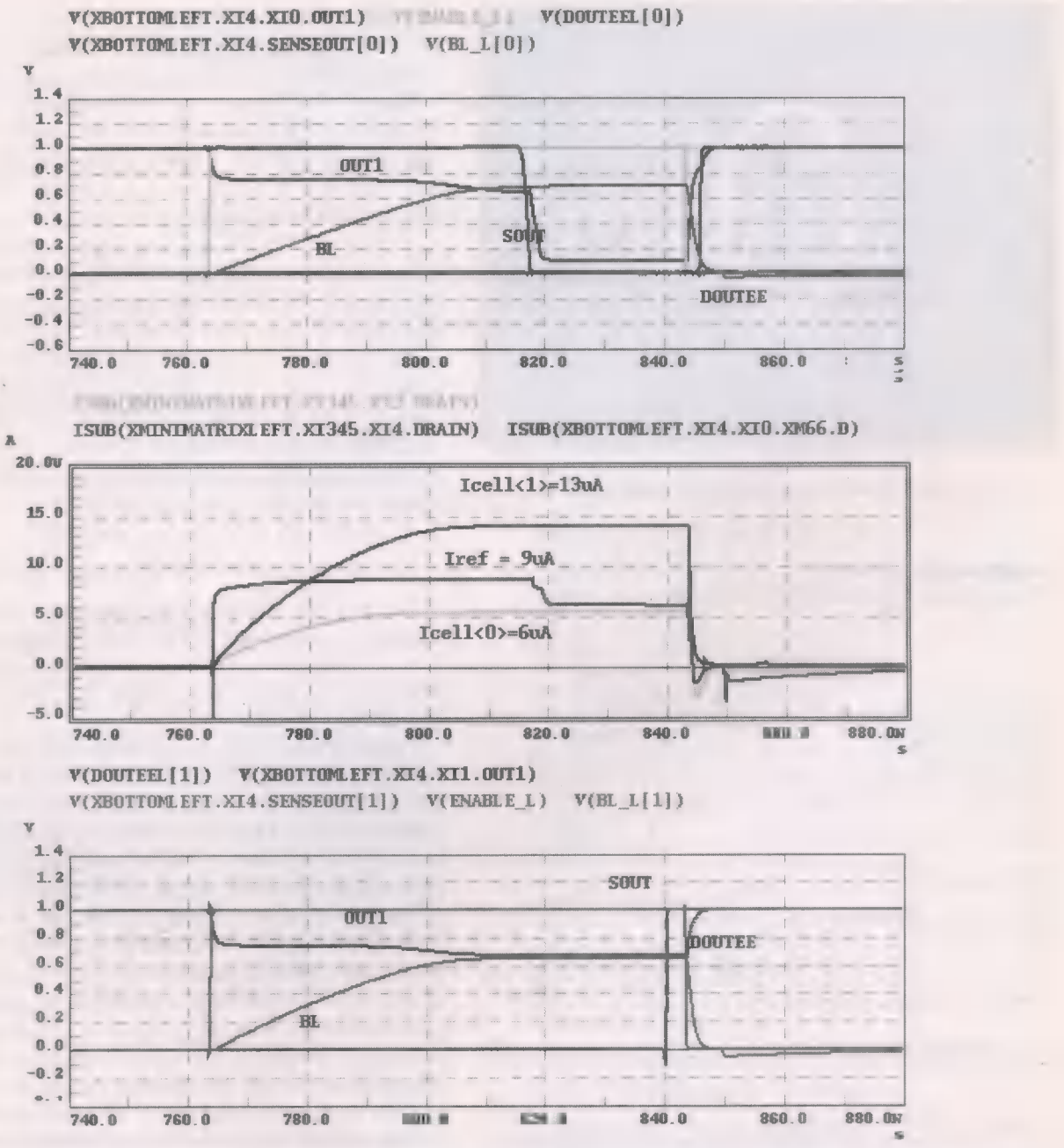


Fig. 7(b). *(Continued.)* Simulation results of the sense amplifier under a 1-V power supply assuming the cell deleted with a cell current equal to 6 μA (upper plot), the cell programmed with a current cell equal to 13 μA (lower plot).

fast pre-charge phase the current used to charge the bitline is equal to the bias current multiplied by the mirror factor formed by the series of M7 and M8 on one side and transistor M3 on the other given by

$$K1 = \frac{(W/L)_3}{(W/L)_{7/8}}. \tag{3}$$

When the bitline reaches an NMOS threshold voltage, transistor M9 begins to sink the current reducing voltage reference V_{REF}, because the drain-source voltage drop of M8 (or M9) is decreased. To obtain design relationships, which relate the steady state voltage reference, V_{REF}, to the transistor aspect ratios, we can also assume M9 is equal to M8 and M7. In the steady state, transistor M9 has the same gate voltage as M8, and we can approximate M8 and M9 with an equivalent transistor of the same length and width equal to twice M8 (or M9).

To further increase speed during the pre-charge phase, a circuit providing an adjoin current only during the pre-charge time slot is added as well (see Fig. 5). When the bitline voltage is lower than an NMOS threshold voltage, the circuit feeds an adjoin current to the bitline node which is equal to the bias current amplified by gain K2 with the two current mirrors M10–M11 and M12–M13 in Fig. 4. After having reached the threshold voltage, transistor M14 switches on and a current mirror between M1 and M14 is created. Then transistor M14 sinks current I_{bias}, which means the circuit M10–M13 is switched off.

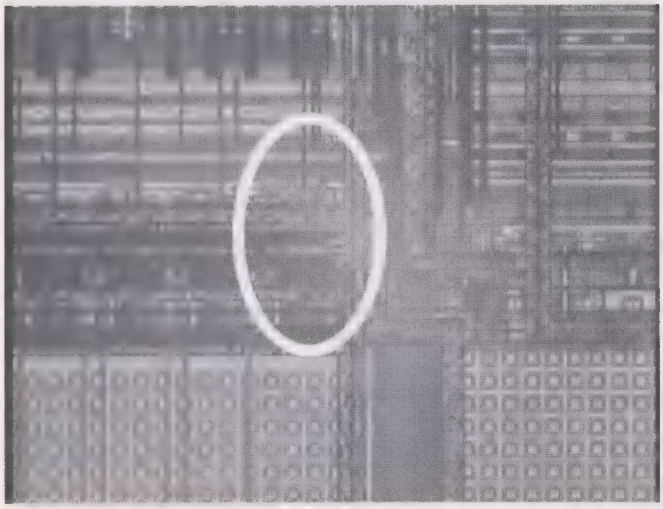


Fig. 8. Sense amplifier microphotograph (inside the bold circle).

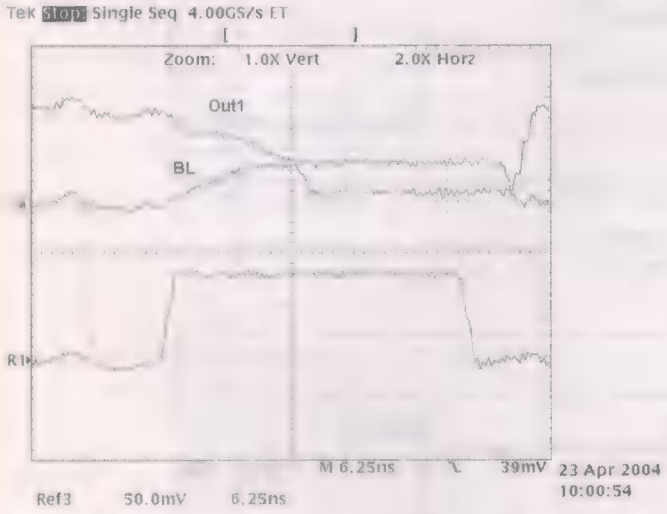


Fig. 9. Experimental results from the sense amplifier read access time on an erased cell.

In conclusion, the pre-charge time is reduced through the two circuits in Figs. 4 and 5, which can be split into two main contributions approximated by

$$t_{\rm pre} = \frac{C_{\rm BL}}{I_{\rm bias}(K1+K2)} V_{\rm TH} + \frac{C_{\rm BL}}{I_B}\left(V_{\rm REF} - V_{\rm TH}\right) \quad (4)$$

where current I_B is given by M3 in saturation with a gate voltage equal to $V_{\rm REF}$.

The complete sense amplifier is shown in Fig. 6. To properly amplify the output voltage, a two stage amplifier is added. The first stage compares the internal output, OUT1, with the voltage on the bitline (i.e., the reference voltage). It is made up of the low-voltage differential amplifier M15–M16 biased with two current generators, $I_{\rm bias}$, which include a folded mirror active load to improve the gain by a factor of two without limiting the minimum allowable power supply [24]. The second stage is the simple inverter M19–M20.

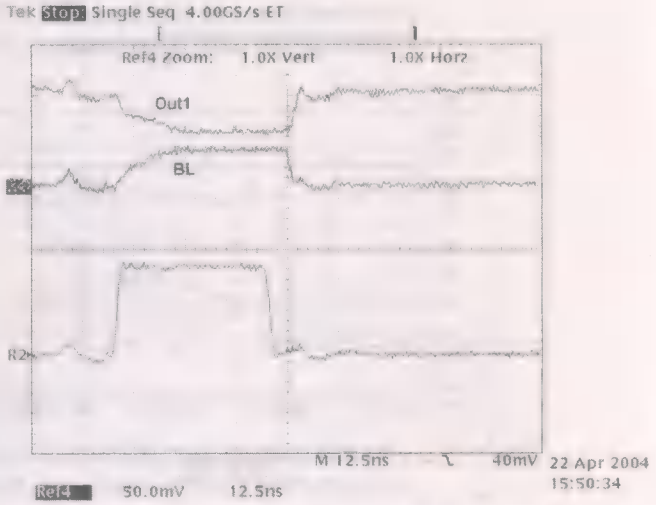


Fig. 10. Experimental results from the sense amplifier read access time on a programmed cell.

IV. Simulation and Experimental Results

The very low-voltage sense amplifier presented in previous sections was integrated in a EEPROM memory fabricated in a 0.18-μm EEPROM technology using the transistor aspect ratios summarized in Table I, bias current, $I_{\rm bias}$, equal to 16 μA, and $V_{\rm ref}$ of about 700 mV.

Transient simulations setting a reference current, $I_{\rm ref}$, equal to 10.5 μA, a capacitive load which modeled all the read paths (i.e., the load due to the memory math and the array of sense amplifiers), equal to 1 pF,[2] and various power supply and memory cell currents were carried out. In particular, those at a nominal power supply of 1.65 V under the two critical cases of a memory cell weakly erased and weakly programmed, were modeled with a cell current equal to 6 μA and 13 μA, respectively. They are plotted in Fig. 7(a). In particular, the upper plot refers to the case of an erased cell with an I_C equal to 6 μA, the lower ones to a programmed cell with an I_C equal to 13 μA, and the middle plot shows the level of cell current both in the erase and in the programmed case and the reference current set to 10.5 μA. It is worth noting that the latter case is the most critical since the cell current is closer to the reference current than the other one. The output data obtained sampling the signal $S_{\rm out}$ is also shown in Fig. 7(a), where it is named DOUTEE.

To highlight the variation on the simulated performance of the sense amplifier, simulation results using different power supplies and temperatures are summarized in Table II. In particular, a read access time lower than 50 ns for a current difference of 1 μA can always be obtained.

In order to show the correct behavior of the sense amplifier with a 1-V power supply, transient simulations setting a reference current, $I_{\rm ref}$, equal to 9 μA, a capacitive load equal to 1 pF, under the two critical cases of a memory cell weakly erased and weakly programmed, are plot in Fig. 7(b). In particular, in

[2]For the capacitive load it is used a typical value. Indeed, for a 64 kB memory we have 512 cells connected to each BL, resulting in an equivalent capacitive load of about 450 fF. The metal interconnection which connect all the cells drain has a parasitic capacitance of about 370 fF. Finally, the bus interconnection between sense amplifier and colum decoder gives a contribute of about 250 fF.

Fig. 7(b) the upper plot refers to the case of an erased cell with an I_C equal to 6 μA, and the lower ones to a programmed cell with an I_C equal to 13 μA. Moreover, in the last row of Table II read access time and minimum current compared for 1-V power supply at 27°C are reported. Of course, at 1-V power supply the access time is increased, but as shown in Fig. 7(b) the sense amplifier behavior is correct.

The sense amplifier has a silicon area of about 600 μm^2 and its microphotograph is shown in Fig. 8.

Experimental results are plotted in Figs. 9 and 10. In particular, Fig. 9 refers to an erased cell (i.e., a cell current lower than the reference current set to 10 μA) and Fig. 10 to a programmed cell. Measurements were carried out on a reading cycle of an erased cell at 1.65 V. The measurements show a correct output level after about 20 ns, allowing a read access time of about 30 ns. Moreover, as expected, measured average current consumption is about 60 μA. More specifically, it is equal to 60 μA if measured in a time window of 100 ns, while it is equal to 76 μA if measured during the read period of about 38 ns.

V. Conclusion

A current sense amplifier solution for nonvolatile memories has been presented. The circuit exhibits good performance over a very low-voltage range, allowing extensive control of both speed and bitline voltage levels, even under the extreme condition of power supplies as low as 1.35 V. Moreover, the absence of any cascoding technique in the bitline pre-charging scheme allows the circuit to function with power supplies as low as 1 V, as a power supply higher than the sum of a threshold voltage and a drain-source saturation is needed.

The sense amplifier was implemented and validated with a 0.18-μm EEPROM technology for Smart Card applications and enables a read access time lower than 30 ns.

Appendix

Using the well-known Shicman–Hodges equation which means neglecting short channel effects, on the circuit in Fig. 3 when $I_C = 0$ the ratios of the drain current of transistor M1 and M3 and that of transistor M5 and M4, I_1/I_3 and I_1/I_3 respectively, is given by

$$\frac{I_1}{I_3} = \frac{(V_{GS1} - V_{Tn})(1 + \lambda_n V_{GS1})}{(V_{REF} - V_{Tn})(1 + \lambda_n V_{DS3})} = \frac{(V_{GS1} - V_{Tn})(1 + \lambda_n V_{GS1})}{(V_{REF} - V_{Tn})[1 + \lambda_n (V_{DD} - V_{SG4})]} \quad \text{(A1)}$$

$$\frac{I_5}{I_4} = \frac{1 + \lambda_p V_{SD5}}{1 + \lambda_p V_{SG4}} = \frac{1 + \lambda_p (V_{DD} - V_{GS1})}{1 + \lambda_p V_{SG4}} \quad \text{(A2)}$$

where V_{Tn} is the threshold voltage of the NMOS transistor, λ_n and λ_p are the channel length modulation parameters of NMOS and PMOS transistors, respectively, and the other parameters have the usual meaning. Since currents I_1 and I_3 are equal to I_5 and I_4, respectively, we get

$$\frac{(V_{GS1} - V_{Tn})(1 + \lambda_n V_{GS1})}{1 + \lambda_p (V_{DD} - V_{GS1})} = \frac{(V_{REF} - V_{Tn})[1 + \lambda_n (V_{DD} - V_{SG4})]}{1 + \lambda_p V_{SG4}} \quad \text{(A3)}$$

Relationship (A3) states that $V_{GS1} = V_{REF}$, neglecting the channel length modulation (i.e., $\lambda_n = \lambda_p = 0$), or matching the source-drain voltages for the NMOS and the PMOS transistor couple. The same results can be achieved by considering short channel effects.

References

[1] A. Sharma, *Semiconductor Memories*. New York: IEEE Press, 1997.

[2] P. Cappelletti, C. Golla, P. Olivo, and E. Zanoni, *Flash Memories*. Boston, MA: Kluwer, 1999.

[3] T. Haraszti, *CMOS Memory Circuits*. Boston, MA: Kluwer, 2001.

[4] A. Matsuzawa, "Low-voltage and low-power circuit design for mixed analog/digital systems in portable equipment," *IEEE J. Solid-State Circuits*, vol. 29, no. 4, pp. 470–486, Apr. 1994.

[5] A. Chandrakasana and R. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, no. 4, pp. 498–523, Apr. 1995.

[6] A. Bellaouar and M. Elmasry, *Low-Power Digital VLSI Design Circuits and Systems*. Boston, MA: Kluwer, 1995.

[7] A. Chandrakasan and R. Brodersen, *Low Power Digital CMOS Design*. Boston, MA: Kluwer, 1995.

[8] J. Rabaey and M. Pedram, *Low Power Design Methodologies*. Boston, MA: Kluwer, 1996.

[9] E. Sanchez-Sinencio and A. Andreou, Eds., *Low-Voltage/Low-Power Integrated Circuits and Systems*. New York: IEEE Press, 1999.

[10] J. Kuo and J. Luo, *Low-Voltage CMOS VLSI Circuits*. New York: Wiley Interscience, 1999.

[11] V. Oklobdzija, Ed., *High-Performance System Design (Circuits and Logic)*. New York: IEEE Press, 1999.

[12] K. Roy and S. Prasad, *Low-Power CMOS VLSI Circuit Design*. New York: Wiley Interscience, 2000.

[13] Y. Miyawaki, "A 29-mm^2 1.8-V-only 16-Mb DINOR flash memory with gate-protected-poly-diode (GPPD) charge pump," *IEEE J. Solid-State Circuits*, vol. 34, no. 11, pp. 1551–1554, Nov. 1999.

[14] N. Otsuka and M. Horowitz, "Circuit techniques for 1.5-V power supply flash memory," *IEEE J. Solid-State Circuits*, vol. 32, no. 8, pp. 1217–1230, Aug. 1997.

[15] T. Tanzawa, Y. Takano, T. Taura, and S. Atsumi, "Design of a sense circuit for low-voltage flash memories," *IEEE J. Solid-State Circuits*, vol. 35, no. 10, pp. 1415–1421, Oct. 2000.

[16] S. Atsumi *et al.*, "A channel-erasing 1.8-V-only 32-Mb NOR flash EEPROM with bitline direct sensing scheme," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1648–1653, Nov. 2000.

[17] T. Tanzawa, Y. Tanaka, K. Takeuchi, and H. Nakamura, "Circuit techniques for a 1.8-V-only NAND flash memory," *IEEE J. Solid-State Circuits*, vol. 37, no. 1, pp. 84–89, Jan. 2002.

[18] W. Rankl and W. Effing, *Smart Card Handbook*, 2nd ed. New York: Wiley, 2000.

[19] P. Rakers, L. Connell, T. Collins, and D. Russell, "Secure contactless smartcard ASIC with PDA protection," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 559–565, Mar. 2001.

[20] A. Abrial, J. Bouvier, M. Renaudin, P. Senn, and P. Vivet, "A new contactless smart card IC using an on-chip antenna and an asynchronous microcontroller," *IEEE J. Solid-State Circuits*, vol. 36, no. 7, pp. 1101–1106, Jul. 2001.

[21] R. Micheloni, M. Crippa, M. Sangalli, and G. Campardo, "The flash memory read path: Building blocks and critical aspects," *Proc. IEEE*, vol. 91, no. 4, pp. 537–553, Apr. 2003.

[22] D. Freitas and K. Current, "CMOS current comparator circuit," *Electron. Lett.*, vol. 19, no. 17, pp. 695–697, Aug. 1983.

[23] G. Palmisano and G. Palumbo, "High performance CMOS current comparator design," *IEEE Trans. Circuits Syst. II*, vol. 43, no. 12, pp. 785–790, Dec. 1996.

[24] G. Palmisano, G. Palumbo, and R. Salerno, "1.5-V high-drive capability CMOS opamp," *IEEE J. Solid-State Circuits*, vol. 34, no. 2, pp. 248–252, Feb. 1999.

Antonino Conte was born in Porto Empedocle (Agrigento), Italy, on May 5, 1966. He received the Degree in electronic engineering (*summa cum laude*) from the University of Palermo, Italy, in 1992.

He worked for Siemens Telecomunicazioni for two years in the field of radio systems. Then he joined STMicroelectronics in 1995 as an Analog Designer working on NVM Memory macrocells for Smart Card application, and in this field he holds more than 14 patents. He participated in the NVSMW 2000, presenting a paper on sense amplifier architecture. He is currently responsible for the Design Development of NVM memory macrocells in the Catania Smart Card Design Center of STMicroelectronics.

Gianbattista Lo Giudice was born in Palermo, Italy, on July 11, 1971. He received the Degree in electronic engineering (*summa cum laude*) from the University of Palermo, Italy, in 1997.

He worked for Pirelli SpA for a year in the Research and Development Department in a group that worked on passive optelectronic devices. He joined STMicroelectronics in 1999 as an Analog Designer working on NVM memory macrocells for Smart Card application and in this field he holds two patents. He is currently responsible for the design of EEPROM macrocells in the Catania Smart Card Design Center of STMicroelectronics.

Gaetano Palumbo (M'91–SM'98) was born in Catania, Italy, in 1964. He received the Laurea degree in electrical engineering in 1988 and the Ph.D. degree from the University of Catania in 1993.

Since 1993, he has been conducting courses on electronic devices, electronics for digital systems and basic electronics. In 1994, he joined the DEES (Dipartimento Elettrico Elettronico e Sistemistico), now DIEES (Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi), at the University of Catania as a Researcher, subsequently becoming Associate Professor in 1998. Since 2000, he has been a full Professor in the same department. His primary research interest has been analog circuits with particular emphasis on feedback circuits, compensation techniques, current-mode approach, and low-voltage circuits. His research has also included digital circuits with emphasis on bipolar and MOS current-mode digital circuits, adiabatic circuits, and high-performance building blocks focused on achieving optimum speed within the constraint of low-power operation. In all these fields, he is developing research activities in collaboration with STMicroelectronics of Catania. He was the co-author of the books *CMOS Current Amplifiers*, *Feedback Amplifiers: Theory and Design* and *Model and Design of Bipolar and MOS Current-Mode Logic (CML, ECL and SCL Digital Circuits)* (Kluwer, 1999, 2002, and 2004). He is a contributor to the *Wiley Encyclopedia of Electrical and Electronics Engineering*. In addition, he is the author or co-author of more than 220 scientific papers in international journals (almost 90) and in conferences. From June 1999 to the end of 2001, he served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–PART I for the topic "Analog Circuits and Filters." In 2004, he served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–PART I for the topic "Digital Circuits and Systems."

Prof. Palumbo received the Darlington Award in 2003.

Alfredo Signorello was born in Catania, Italy, on May 17, 1971. He received the Laurea degree in electronics engineering from the University of Catania in 2001.

He joined STMicroelectronics in 2001 working on NVM memory macrocells for Smart Card application as an Analog Designer, and he holds three patents in this field. His current interests are the development of macrocells for the nonvolatile memory.

A Novel High-Speed Sense Amplifier for Bi-NOR Flash Memories

Chiu-Chiao Chung, Hongchin Lin, *Member, IEEE*, and Yen-Tai Lin

***Abstract*—A novel high-speed current-mode sense amplifier is proposed for Bi-NOR flash memory designs. Program and erasure of the Bi-NOR technologies employ bi-directional channel FN tunneling with localized shallow P-well structures to realize the high-reliability, high-speed, and low-power operation. The proposed sensing circuit with advanced cross-coupled structure by connecting the gates of clamping transistors to the cross-coupled nodes provides excellent immunity against mismatch compared with the other sense amplifiers. Furthermore, the sensing times for various current differences and bitline capacitances and resistances are all superior to the others. The agreement between simulation and measurement indicates the sensing speed reaches 2 ns for the threshold voltage difference of lower than 1 V at 1.8-V supply voltage even with the high threshold voltage of the peripheral CMOS transistors up to 0.8 V.**

***Index Terms*—Advanced cross-couple, Bi-NOR, clamping transistor, flash memory, FN tunneling, mismatch, threshold voltage.**

I. Introduction

FOR contemporary memories, array structures and periphery circuits, such as decoders, charge pumps, level shifters, and sense amplifiers, determine the overall system performance in terms of power dissipation and access speed. The high-speed low-power sense amplifier is one of the critical components. Due to low-voltage operation, current sensing techniques have received a lot of attention in the last decade. Many sense amplifiers based on cross-coupled transistor structures were designed to overcome the loading effects [1]–[3] for DRAM or SRAM, but few have been discussed about the mismatch of sense amplifiers. Another category of memories is flash memory [4], [5]. The trend is not only high-density and low-voltage, but also multi-level. Therefore, the threshold voltage deviation of the programmed memory cells has to be well controlled for low-voltage operation. The sense amplifiers require high sensitivity and excellent mismatch immunity in threshold voltage and W/L (channel width/channel length) ratio of devices.

For flash memories, comparison of current difference between the flash cell and the reference cell is the direct and fast method to read the data. However, for the Bi-NOR [6], [7] flash memory arrays, most of the sensing circuits developed for the conventional flash memory cells [8], [9], such as the simple four-transistor sense amplifier [10], PMOS bias type sense amplifier [11], and differential latch type sense amplifier [12], are not appropriate. Since these sense amplifiers were designed for draining cell current at the drain node of the flash cell, their bitlines were usually pre-charged to high before sensing. However, the current direction for Bi-NOR cells is reversed. The sense amplifier drains the current of the flash cell at the source node, thus the bias at the bitline source node has to be low enough for the cell current flowing to the sense amplifier. Though the clamped bitline (CBL) sense amplifier [13] was appropriate for the Bi-NOR cells, it would result in higher power consumption, lower sensing speed, and poor mismatch effects due to the equalization of the bitlines before sensing.

To comply with these restrictions, we propose a new sense amplifier (NSA) that utilizes advanced cross-coupled structure by connecting the gates of the clamping MOS transistors to the cross-coupled nodes to improve the mismatch characteristics and reduce the power consumption without scarification of sensing time. The mismatch is also improved if the equalization between the drains of the two clamping MOS transistors is removed, since the currents from the selected cell and the reference cell slightly charge the drains before sensing.

The new circuit and its operation principle for Bi-NOR cells are described in Section II. Section III compares the sensing speed versus threshold voltage difference, bitline capacitance, and channel length mismatch with the clamped bitline sensing scheme. The theory of mismatch improvement is also given in this section. In Section IV, the measurement results show the agreement with simulations. Section V is the conclusion.

II. The New Sense Amplifier and Its Operation

The flash memory cell used in this study is based on the Bi-NOR technology [6], [7], which uses bi-directional channel FN tunneling with localized shallow P-well structure to realize the high-reliability, high-speed, and low-power operation. The conduction channel width of the flash cell is no longer one-dimensional. Fig. 1(a) illustrates the cross-sectional view of Bi-NOR flash memory cells. The current consists of the conventional current path (solid arrow) and the side conduction path shown by the dashed arrow. Since the electron current is flowing from the width, length, and bottom (deep N-well) directions, more than 15% read conduction current enhances the read performance. The typical operating conditions for Bi-NOR cell are listed in Table I. Fig. 1(b) shows the read path from an array to the sense amplifier. For a selected cell, since the drains of the flash cells in the same row are connected and biased at 1 V from the source switch, the current has to flow to the sense amplifier at the bitline of the flash cell. Therefore, the bias at the bitline must be close to zero to comply with

Manuscript received March 4, 2004; revised August 1, 2004. This work was supported by NSC of Taiwan, R.O.C., under NSC91-2622-E-007-033.

C.-C. Chung and H. Lin are with the Department of Electrical Engineering, National Chung-Hsing University, Taichung 402, Taiwan, R.O.C. (e-mail: hclin@dragon.nchu.edu.tw).

Y.-T. Lin is with eMemory Technology Inc., Hsinchu 300, Taiwan, R.O.C.

Digital Object Identifier 10.1109/JSSC.2004.840965

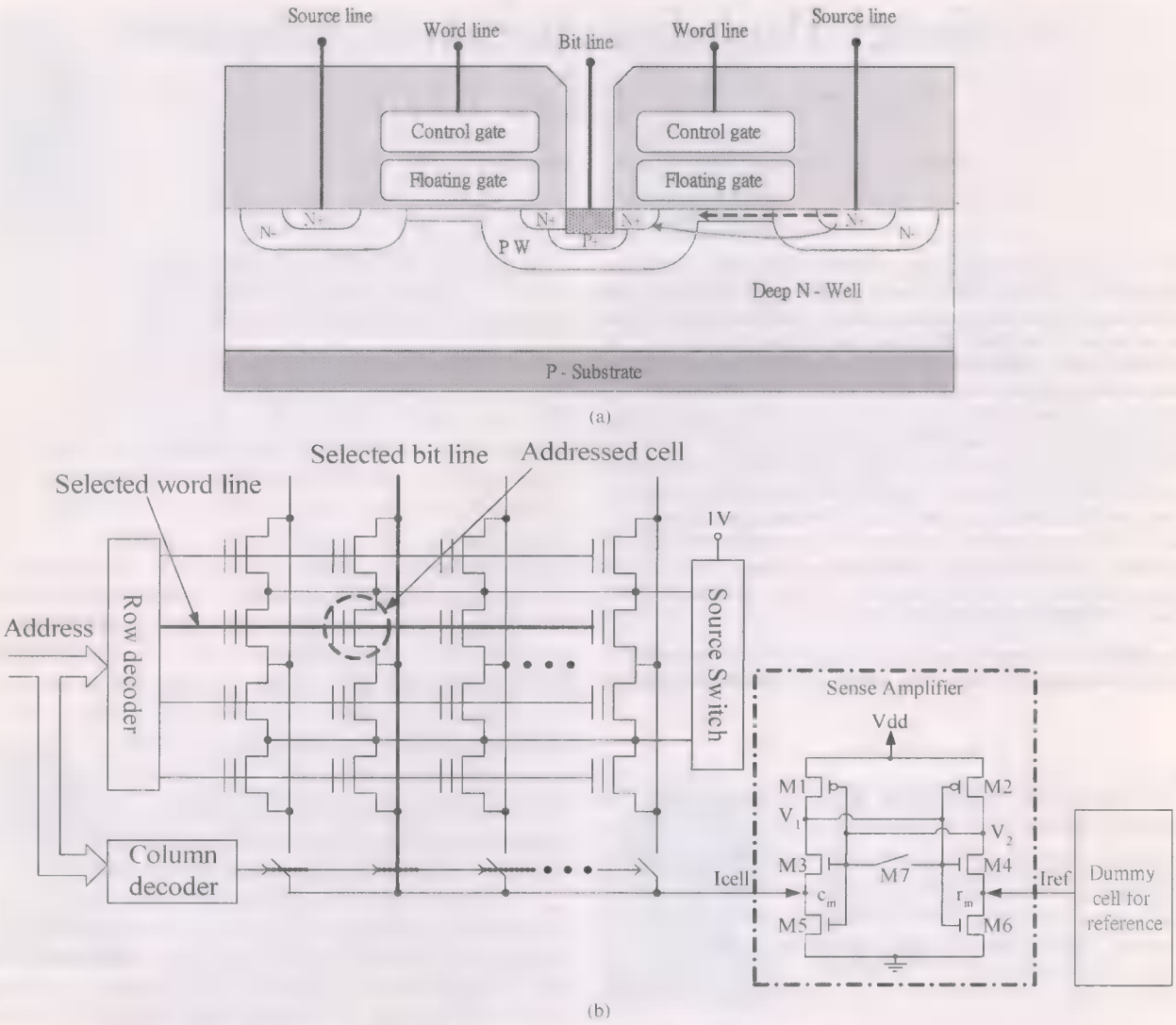


Fig. 1. (a) Cross-sectional view of the Bi-NOR flash memory cell. (b) The read path in an array organization.

TABLE I
TYPICAL OPERATING CONDITIONS FOR THE BI-NOR CELL

Operation	Bit line		Word line		Source line	Deep N-Well
	Select	unselect	select	unselect		
Program	6V	0V	-10V	0V	Float	Float
Erase	Float	Float	18V	0V	0V	0V
Read	0V	1V	4V	0V	1V	1V

the requirement. This new operation makes most of the sense amplifiers designed for the conventional flash cell arrays not appropriate for the new cell array.

Generally, the sensing circuit is composed of a current source transporting the cell's contents through the bitline to the data line, and a latch stage converting the differential current in the data line to the output node. According to the Bi-NOR cell array mentioned above, the new current-mode sense amplifier shown in Fig. 2(a) employs the cross-coupled latch structure (M1–M4) with sensor activation (Men) and equalization of output nodes (M7). Transistors M5 and M6 clamp the bitline voltage close to ground, and the sensing nodes (c_{in} and r_{in}) drain currents from the selected cell and the reference cell, respectively. The $R_{bitline}$ and $C_{bitline}$ represent the parasitic resistance and capacitance at the bitline. The timing diagram of signals SE, En, Nodes a, b, and out for the new sense amplifier is illustrated in Fig. 2(b).

The operation of the sense amplifier can be divided into three phases: pre-charge, signal amplification, and reset for the next operation. In the pre-charge phase, the appropriate signals are applied to force the sensing nodes to certain potentials. In the amplification phase, the comparison and amplification are executed between the sensing nodes, so the content of the selected memory cell is retrieved. After that, the sense amplifier is reset for the next operation.

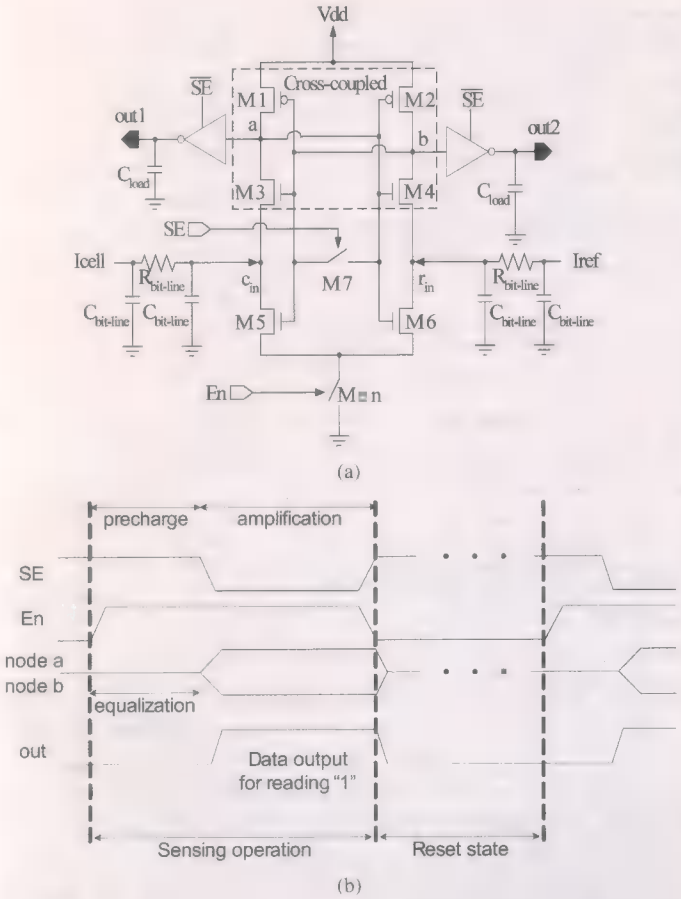


Fig. 2. (a) Circuit diagram of the new sense amplifier. (b) Timing diagram for the operation of the new sense amplifier.

The sensing operation starts by turning on Switch Men and Switch M7. During the pre-charge phase, the output node voltages are equalized ($V\mathrm{a}{=}V\mathrm{b}$) so that the currents in M1 and M2 are the same. For the case of $I_{\mathrm{cell}} > I_{\mathrm{ref}}$, the current through M5 will be larger than that of M6 ($I_{\mathrm{M5}} > I_{\mathrm{M6}}$). Therefore, the bias at Node $\mathrm{c_{in}}$ is slightly higher than that at $\mathrm{r_{in}}$. In the meanwhile, since M3 and M4 are all in the saturation region, the gate to source voltage (V_{gs}) of M3 is less than that of M4 ($V_{gs3} < V_{gs4}$), the current through M3 is smaller than that of M4 ($I_{\mathrm{M3}} < I_{\mathrm{M4}}$).

At the end of the pre-charge cycle, M7 is turned off, so transistors M1–M4 act as a high-gain positive feedback amplifier. Due to positive feedback, the impedance looking into the source node of either M3 or M4 is negative. That makes M3 and M4 begin to source the currents when M7 is turned off. Since M4 has stronger ability than M3 does to discharge the voltage at the node b, the different currents flowing through the drains of transistors M3 and M4 amplify the voltage difference across the output nodes (a and b) of the sensing amplifier. During the pre-charge phase, it is important that the sizes of clamping transistors M5 and M6 should be chosen slightly larger to allow them biased in linear region, thus activating the regeneration procedure of inverter pairs (M1/M3 and M2/M4) as a latch in the later amplification phase.

Since the inputs of the sense amplifier are low-impedance current sensing nodes, the high capacitive bitlines only need to be charged slightly for sensing operation. This results in the minimal influence of the sensing speed for various bitline capacitances and current differences. In addition, due to the fact that the potentials of bitlines always keep low during the sensing operation, power consumption is significantly reduced. Another important feature is improvement of the mismatch problem, which will be explained in the next section.

III. Performance Evaluation

The new sensing circuit was designed and fabricated using 0.25-μm Bi-NOR flash memory technologies with 0.4-μm CMOS transistors with threshold voltage $|V_t| \cong 0.8$ V for peripheral circuits at supply voltage of 1.8 V. Fig. 3(a) shows the simulated waveforms of the signal SE, the nodes a, b, and out of the proposed sense amplifier in the case of $I_{\mathrm{cell}} > I_{\mathrm{ref}}$ with the output load capacitance of 20 fF. The simulation results show the sensing speed is about 2.3 ns for the current difference ($\Delta I = I_{\mathrm{cell}} - I_{\mathrm{ref}}$) of 6.5 μA. Fig. 3(b) gives the waveforms of current input nodes with bitline resistance of 320 Ω and capacitance of 2 pF for the flash cell ($\mathrm{c_{in}}$) and the reference cell ($\mathrm{r_{in}}$). As mentioned before, the potentials at the sources of the cells are pretty low. They are pre-charged to 0.5 V at $\mathrm{c_{in}}$ and 0.1 V at $\mathrm{r_{in}}$ for the pre-charging time of 30 ns.

In order to evaluate the proposed design, the clamped bitline (CBL) sense amplifier [13] illustrated in Fig. 4(a) is compared. Its small-signal equivalent model for the typical cross-coupled circuit is given in Fig. 4(b), in which Cd is the equivalent capacitance at the output nodes of the sense amplifier including the Miller capacitance from the diffusion capacitances $\mathrm{C_{gd}}$, and Rd includes the parallel combination of the output resistances of both n-channel and p-channel transistors. The clamp transistors M5 and M6 are biased in the linear region with equivalent capacitances Cs and conductance g_{ds}. Resistors $\mathrm{R_{12}}$ and $\mathrm{R_{34}}$ mimic the small impedances of switches during the equalization phase. For the current difference $\Delta I = I_{\mathrm{cell}} - I_{\mathrm{ref}} > 0$, the voltage difference between Nodes $\mathrm{c_{in}}$ and $\mathrm{r_{in}}$ is defined as $\Delta V = V_3 - V_4$. For the CBL sense amplifier, ΔV_{CBL} is

$$\Delta V_{CBL} \cong \frac{1}{g_{ds}} \left[(I_{\mathrm{cell}} + I_{M3} - \Delta I_{34}) - (I_{\mathrm{ref}} + I_{M4} + \Delta I_{34}) \right] \quad (1)$$

where g_{ds} is the drain-source conductance of M5 and M6 and ΔI_{34} is the current through $\mathrm{R_{34}}$.

On the other hand, the new sense amplifier with the gates M5 and M6 are connected to the cross-coupled nodes, and the currents through the flash cell, M3, and M4 are denoted as I'_{cell}, I'_{M3}, and I'_{M4}, respectively. Therefore, the voltage difference of proposed circuit ΔV_{NSA} between $\mathrm{c_{in}}$ and $\mathrm{r_{in}}$ without ΔI_{34} term becomes

$$\Delta V_{NSA} \cong \frac{1}{g_{ds}} \left[(I'_{\mathrm{cell}} + I'_{M3}) - (I_{\mathrm{ref}} + I'_{M4}) \right]. \quad (2)$$

For the same sense amplification capability, $\Delta V_{CBL} = \Delta V_{NSA}$, (1) should be equal to (2):

$$\lfloor (I_{\mathrm{cell}} + I_{M3} - \Delta I_{34}) - (I_{\mathrm{ref}} + I_{M4} + \Delta I_{34}) \rfloor \cong \lfloor (I'_{\mathrm{cell}} + I'_{M3}) - (I'_{\mathrm{ref}} + I'_{M4}) \rfloor. \quad (3)$$

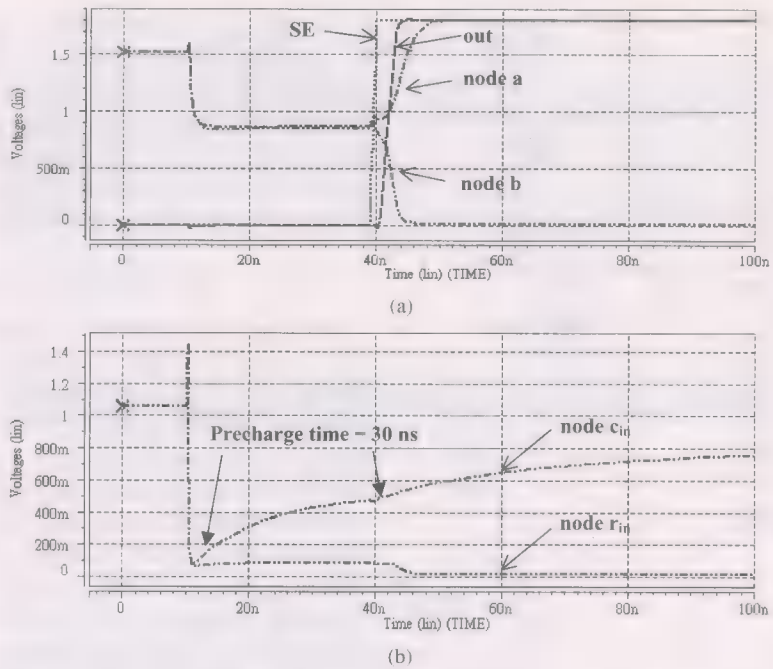


Fig. 3. (a) Simulated waveforms of Signal SE, out, Nodes a and b of the new sense amplifier. (b) Simulated waveforms of Nodes c_{in}, and r_{in} of the new sense amplifier.

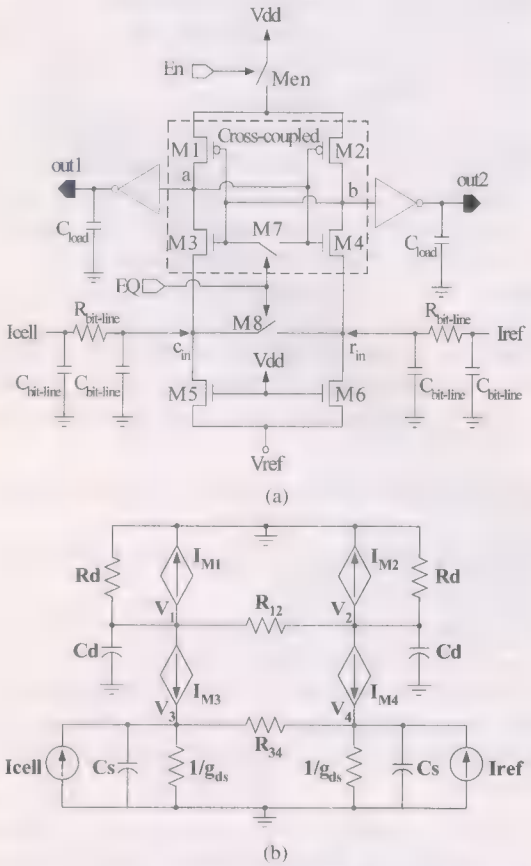


Fig. 4. (a) Circuit diagram of the clamped bitline sense amplifier. (b) Equivalent circuit with M5/M6 denoted as resistors of $1/g_{ds}$.

If we assume $I_{M3} - I_{M4} = I'_{M3} - I'_{M4}$ before the amplifier, it means

$$(I_{\text{cell}} - I_{\text{ref}}) - 2(\Delta I_{34}) \cong (I'_{\text{cell}} - I_{\text{ref}}). \quad (4)$$

It clearly shows that the CBL sense amplifier requires more current difference to compensate the offset [14], since $\Delta I_{34} > 0$ due to $\Delta I = I_{\text{cell}} - I_{\text{ref}} > 0$. The basic difference of the proposed and the CBL sense amplifiers relies on the fact that the equalization device of the proposed circuit is not placed in the current path during the pre-charge phase. Thus, the proposed circuit provides faster response time and better mismatch immunity than the CBL sense amplifier.

The following comparisons were carried out with the same fan-in and fan-out conditions for both circuits with the transistor sizes listed in Table II. Fig. 5 compares the sensing speed and average power dissipation as functions of the current difference for given bitline resistance of 320 Ω and capacitance of 2 pF at $V_{\text{dd}} = 1.8$ V and switch frequency of 25 MHz. The simulations were performed for the current difference of the flash cell (I_{cell}) and the reference cell (I_{ref}) equal to $3 \sim 10$ μA. As expected, the more current difference results in the faster sensing speed. It is obvious that the proposed circuit provides much faster sensing speed and less power consumption compared to the CBL sensing circuits. The reason is that the proposed sense amplifier does not consume sensing current of the cells to either compensate the current path (ΔI_{34}) offset or maintain low biases at c_{in} and r_{in}, thus incurs less power dissipation.

The comparison of sensing speed versus bitline capacitance between the proposed and the CBL sense amplifier for the typical, best and worst transistor models with current difference of 10 μA at $V_{\text{dd}} = 1.8$ V is illustrated in Fig. 6. According to the simulations both sense amplifiers exhibit almost constant sensing delay independent of the bitline load capacitance, since both amplifiers separate the outputs and the bitlines. However, the new circuit has variation of 14% between the typical and the best/worst cases, while the CBL has variation of 22%. The sensing time as functions of pre-charging time for variations in the capacitance and resistance of the bitlines in the memory cell

TABLE II
TRANSISTOR W/L SIZES FOR THE NEW AND THE CBL SENSE AMPLIFIERS

Transistor	NSA	CBL
M1, M2	2 μ / 0.55 μ	
M3, M4	8 μ / 0.55 μ	
M5, M6	25 μ / 0.65 μ	
M7, Men	25 μ / 0.55 μ	
M8	NA	25 μ / 0.55 μ

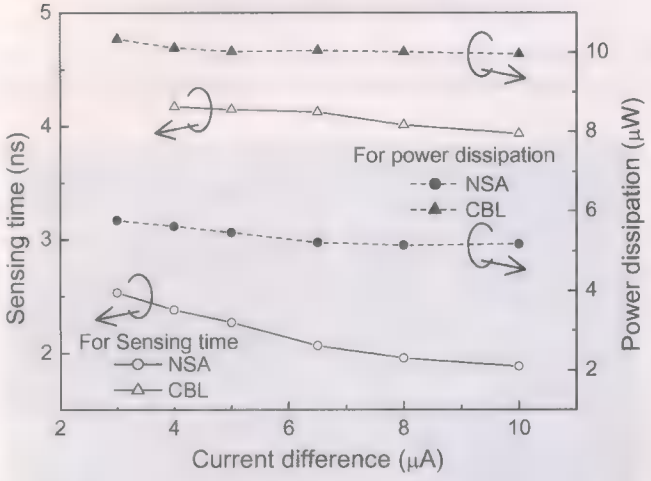


Fig. 5. Simulated sensing speed and average power dissipation for various current differences (ΔI).

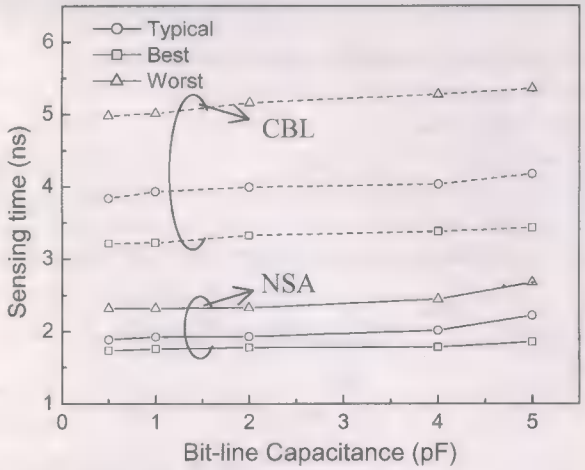


Fig. 6. Sensing speed versus bitline capacitance for different process corners for bitline resistance of 320 Ω.

array is plotted in Fig. 7. In general, the shorter pre-charging time takes the longer sensing time. It can be observed that the pre-charging time is longer with heavier capacitance. However, the variation is not large. Note that the sensing time is barely affected by the resistance variation.

The mismatch in W/L ratio or threshold voltage plays a critical role in the symmetric cross-couple sense amplifiers, since it may result in erroneous sensing output. A simplified model shown in Fig. 8 explains the effect of mismatch in the sensing

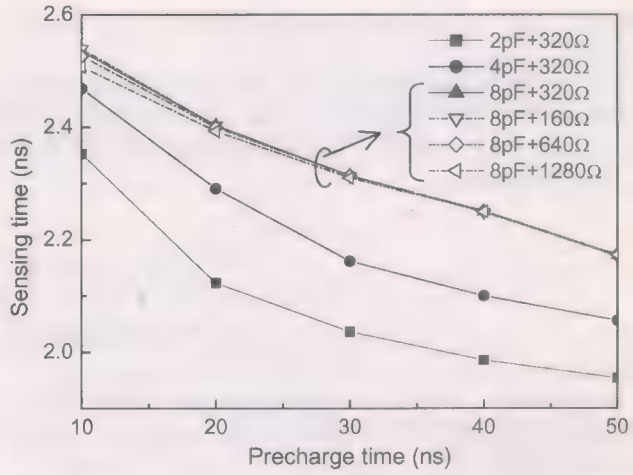


Fig. 7. Sensing speed versus pre-charging time with respect to various bitline resistance and capacitance.

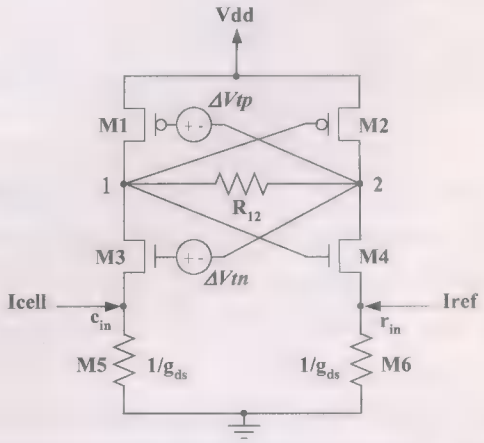


Fig. 8. Equivalent circuit of the new sense amplifier with threshold voltage mismatches.

operation. The ΔV_{tp} and ΔV_{tn} represent the threshold voltage mismatch of PMOS and NMOS transistors, respectively, while g_{ds} denotes as the identical drain to source channel conductance of M5 and M6. By assuming no mismatch of M5 and M6 in the following analysis, the worst polarity for the offset voltage in threshold voltage at the regenerative nodes (Nodes 1 and 2) may be expressed as

$$V_{\text{offset}} = \Delta V_{tn} + \Delta V_{tp} = (g_{mn}\Delta V_{tn} + g_{mp}\Delta V_{tp}) \cdot \mathrm{R}_{12} \quad (5)$$

where g_{mn} and g_{mp} are the transconductances of PMOS and NMOS transistors, and the offset voltage in threshold voltage mismatch is translated into a current mismatch at the drain with a gain of g_m through resistance R_{12}. Since the current difference between the selected cell and reference cell $\Delta I = I_{\text{cell}} - I_{\text{ref}}$, which results in a differential voltage V_{diff} representing the data of selected cell to be read. V_{diff} can be written as

$$V_{\text{diff}} = \Delta I \cdot R_{12}. \quad (6)$$

The ratio of the differential voltage across the differential nodes to the offset voltage called safety margin is defined as [15]

$$\text{Margin} = \frac{V_{\text{diff}}}{V_{\text{offset}}} = \frac{\Delta I}{g_{mn}\Delta V_{tn} + g_{mp}\Delta V_{tp}} \equiv \frac{\Delta I}{I_{\text{offset}}} \quad (7)$$

where I_{offset} is effective offset current, which equals to $g_{mn}\Delta V_{tn} + g_{mp}\Delta V_{tp}$.

The safety margin depends on the transconductance and threshold voltage mismatch of the cross-coupled devices. When switch M7 is on, the currents through M3 and M4 can be approximated as

$$I_{\text{M3}} \cong \frac{\mu_n C_{\text{ox}} W}{2L}(V_{gs3} - V_{tn})^2 = I_{\text{M4}} \cong \frac{\mu_n C_{\text{ox}} W}{2L}(V_{gs4} - V_{tn})^2 . \tag{8}$$

where μ_n is the electron mobility, C_{ox} is the gate capacitance, and V_{tn} is the threshold voltage of NMOS.

In the case of threshold voltage mismatch shown in Fig. 8, the current through M3 is denoted as $I_{\text{M3(mismatch)}}$ varied by a mismatch ΔV_{tn}

$$I_{\text{M3(mismatch)}} \cong \frac{\mu_n C_{\text{ox}} W}{2L}(V_{gs3} - V_{tn} + \Delta V_{tn})^2 . \tag{9}$$

For $I_{\text{cell}} > I_{\text{ref}}$, the source of M3 is charged by a voltage on sensing node c_{in} denoted as V_{cin}, therefore (9) can be rewritten as

$$\begin{aligned} I_{\text{M3(mismatch)}} &\cong \frac{\mu_n C_{\text{ox}} W}{2L}[V_{g3} - (V_{s3} + V_{cin}) - V_{tn} + \Delta V_{tn}]^2 \\ &\cong \frac{\mu_n C_{\text{ox}} W}{2L}[V_{g3} - V_{s3} - V_{tn} + (\Delta V_{tn} - V_{cin})]^2 \\ &\cong \frac{\mu_n C_{\text{ox}} W}{2L}[V_{gs3} - V_{tn} + (\Delta V_{tn} - V_{cin})]^2 \end{aligned} \tag{10}$$

where the V_{g3} and V_{s3} are gate and source voltages of M3, respectively. The threshold voltage mismatch for the proposed circuit is reduced due to the term $(\Delta V_{tn} - V_{cin}) \equiv \Delta V_{tn(NSA)}$ in (10). According to the safety margin definition in (7), $\Delta I/\Delta I_{\text{offset}}$, either the more current difference ΔI or the less offset current benefits the sensing operation in case of mismatch arising. The proposed circuit charges the sensing node c_{in} to reduce the offset current ΔI_{offset} with the term of $(\Delta V_{tn} - V_{cin})$ instead of ΔV_{tn} in (10). However, the CBL sense amplifier does not have this effect due to equalization between c_{in} and r_{in}. Therefore, with the same current difference for amplification, the proposed circuit is superior to the CBL sense amplifier for mismatch improvement.

Since the threshold voltage mismatch can be equivalent to the geometry (W/L ratio) mismatch [15], the worst-case mismatch may be obtained by tuning the possible worse cases at the same time. Therefore, the sensing circuits were simulated using the center dimensions given in Table II with channel length mismatches on M1, M4, and M6, which were selected as $L_{\text{M1}} = L_{\text{M2}} + \Delta L$, $L_{\text{M4}} = L_{\text{M3}} + \Delta L$, and $L_{\text{M6}} = L_{\text{M5}} + \Delta L$, respectively, where ΔL is the channel length mismatch. The sensing speed slightly degrades with channel length mismatch up to $\Delta L = 0.05\ \mu$m for the new sensing circuit, while the CBL sense amplifier cannot afford mismatches beyond $0.015\ \mu$m in case of current difference $\Delta I = 10\ \mu$A at the pre-charging time of 50 ns, as shows in Fig. 9. On the contrary, for the case of $I_{\text{cell}} < I_{\text{ref}}$, the mismatch seems not critical, since the mismatch helps the sensing operation.

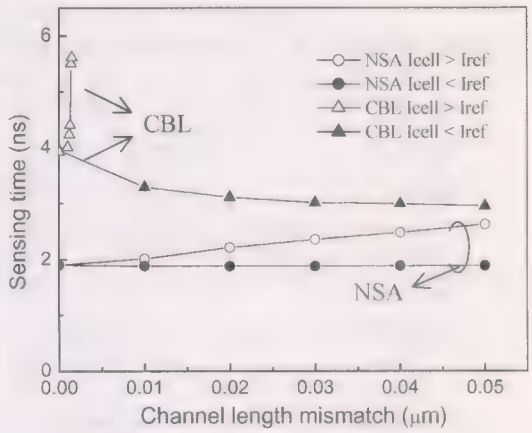


Fig. 9. Sensing speed versus channel length mismatch for $I_{\text{cell}} > I_{\text{ref}}$ and $I_{\text{cell}} < I_{\text{ref}}$ for current difference of $10\ \mu$A, and pre-charging time of 50 ns.

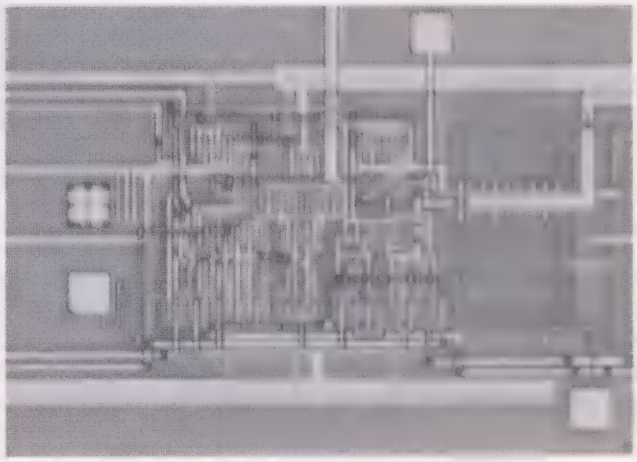


Fig. 10. Chip microphotograph of the new sense amplifier.

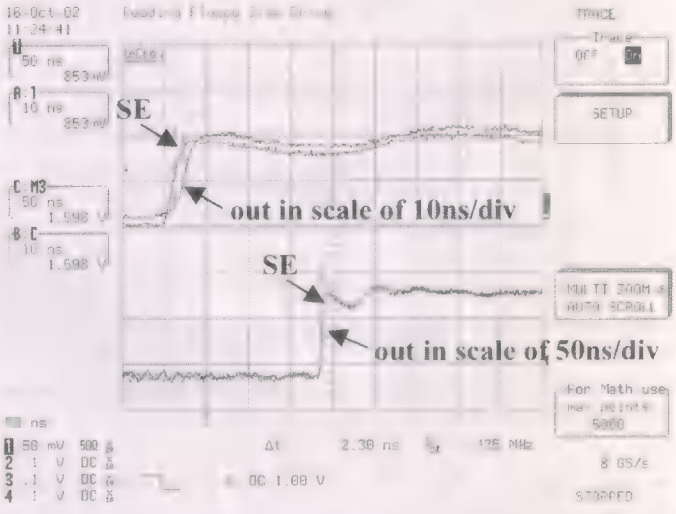


Fig. 11. Measured delay time between the signal SE and node out.

IV. Experimental Results

The chip microphotograph of the new circuit fabricated using 0.25-μm Bi-NOR flash memory with 0.4-μm CMOS for peripheral circuits is presented in Fig. 10. The test chip was designed

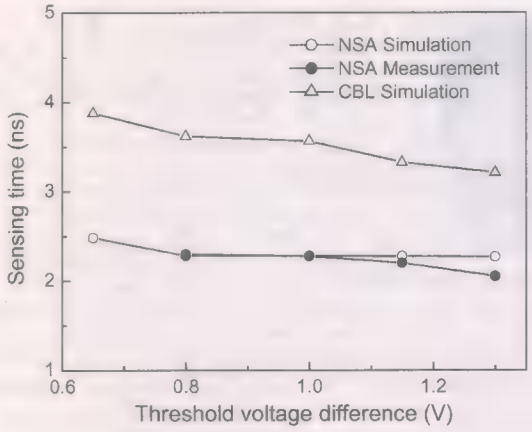


Fig. 12. Sensing speed versus various threshold voltage differences (ΔV).

using the currents generated from the selected cell and reference cell. Each has resistor 320 Ω and two parallel capacitors of 2 pF in between to mimic the parasitic effects in the memory arrays. The cell currents are obtained by applying 1 V to the drains of the selected cell and reference cell with different wordline voltages to the gates of the cells. Since the wordline voltage difference between the selected cell and the reference cell was assumed to be equivalent to the threshold voltage differences between them, the current difference resulted from varying the wordline voltage of the reference cell. Fig. 11 demonstrates that the on-chip measured delay time between the signals SE and output node for the new sense amplifier is about 2.3 ns when the threshold voltage difference is 0.8 V.

The comparison of the sensing delay times between simulation and measurement for the given threshold voltage difference from 0.8 to 1.3 V is shown in Fig. 12. The CBL sense amplifier needs more current difference to compensate the offset, so it takes longer sensing time. The new sense amplifier with the currents slightly charging the sensing nodes before sensing makes the response time shorter. The agreement between measurement and simulation is also observed.

V. Conclusion

A new low-power sensing circuit for 0.25-μm Bi-NOR flash memory technology was designed and measured. The proposed scheme presents outstanding performance with sensing speed reaches 2 ns and power consumption less than 6 μW at switch frequency of 25 MHz and supply voltage of 1.8 V. With the special connection of the gates to the cross-coupled output nodes, the immunity to device mismatch is improved significantly. That also makes the new current-mode sense amplifier much easier to design and fabricate. According to these analyses, it has also proven that the sensing delay of the new sense amplifier is almost independent of the bitline capacitance, which indicates that it is an excellent candidate for higher density memory.

Acknowledgment

The authors would like to acknowledge Power Semiconductor Corporation and eMemory Inc. for their support in chip fabrication and measurement, respectively.

References

[1] J.-S. Wang and H.-Y. Lee, "A new current-mode sense amplifier for low-voltage low-power SRAM design," in *Proc. IEEE Int. ASIC Conf.*, Sep. 1998, pp. 163–167.

[2] S.-M. Yoo *et al.*, "New current-mode sense amplifier for high density DRAM and PIM architectures," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, vol. 4, May 2001, pp. 938–941.

[3] S. M. Wang and C. Y. Wu, "Full current-mode techniques for high-speed CMOS SRAMs," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, vol. 4, May 2002, pp. 580–582.

[4] H. Onoda *et al.*, "A novel cell structure suitable for a 3-V operation sector erase flash memory," in *IEDM Tech. Dig.*, Dec. 1992, pp. 599–602.

[5] H. Kume *et al.*, "A 1.28 μm^2 contactless memory cell technology for a 3 V only 64 M bit EEPROM," in *IEDM Tech. Dig.*, Dec. 1992, pp. 991–993.

[6] C. -S. E. Yang, C. -J. Liu, T. -S. Chao, M. -C. Liaw, and C. -H. C. Hsu, "Novel bi-directional tunneling NOR (Bi-NOR) type 3-D flash memory cell," in *Symp. VLSI Tech. Dig.*, 1999, pp. 85–86.

[7] H.-F. A. Chou *et al.*, "Comprehensive study on a novel bidirectional tunneling program/erase NOR-type (BiNOR) 3-D flash memory cell," *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1386–1393, Jul. 2001.

[8] C. Calligaro, P. Rolandi, N. Telecco, and G. Torelli, "A current-mode sense amplifier for low voltage nonvolatile memories," in *Innovative System in Silicon Conf. Proc.*, 1996, pp. 141–147.

[9] A. Chrisanthopoulos, Y. Moisiadis, A. Varagis, Y. Tsiatouhas, and A. Arapoyanni, "A new flash memory sense amplifier in 0.18 μm CMOS technology," in *Proc. IEEE Int. Conf. Electronics, Circuits, and Systems (ICECS)*, vol. 2, Sep. 2001, pp. 941–944.

[10] E. Seevinck, P. J. Van Beers, and H. Ontrop, "Current mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 525–536, Apr. 1991.

[11] K. Sasaki *et al.*, "A 7-ns 140-mW 1-Mb CMOS SRAM with current sense amplifier," *IEEE J. Solid-State Circuits*, vol. 27, no. 11, pp. 1511–1518, Nov. 1992.

[12] T. Seki *et al.*, "A 6-ns 1-Mb CMOS SRAM with latched sense amplifier," *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 478–483, Apr. 1993.

[13] T. N. Blalock and R. C. Jaeger, "A high-speed clamped bitline current-mode sense amplifier," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 542–548, Apr. 1991.

[14] H. Lin and F. Liang, "A high speed current-mode multi-level identifying circuit for flash memories," *IEICE Trans. Electron.*, vol. E86-C, no. 2, pp. 229–235, 2003.

[15] A. Hajimiri and R. Heald, "Design issues in cross-coupled inverter sense amplifier," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, vol. 2, 1998, pp. 149–152.

Chiu-Chiao Chung was born in Taiwan, R.O.C. She received the B.S. degree in electronic engineering from the Tam-Kang University, Taipei County, Taiwan, in 1983, and the M.S. degree in electrical engineering from the University of Texas, El Paso, in 1989. She is currently pursuing the Ph.D. degree in the Department of Electrical Engineering, National Chung-Hsing University, Taichung, Taiwan.

She joined the Nan-Kai College, Nan-Tao County, Taiwan, in August 1990 as ■ Lecturer in the Department of Electrical Engineering. Her research involves memory circuit design, and Flash memory technology and device design.

Hongchin Lin (M'87) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1986 and the M.S. and Ph.D. degrees from the University of Maryland, College Park, in 1989 and 1992, respectively.

From 1992 to 1995, he was with Integrated Technology Division, Advanced Micro Devices, Sunnyvale, CA. In 1995, he joined the Department of Electrical Engineering, National Chung-Hsing University, Taichung, Taiwan, and was promoted to a full Professor in 2003. His current research interests include VLSI circuit design, semiconductor memory devices and circuits, and VLSI implementation of wireless communication systems.

Yuan-Tai Lin was born in Taiwan, R.O.C. He received the B.S. and M.S. degrees in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 1981 and 1983, respectively.

He joined the ERSO (Electronic Research and Service Organization) of ITRI (Industrial Technology Research Institute) for SRAM/DRAM circuit design when he graduated from NTHU. From 1996 to 1998, he was with Vanguard International Semiconductor Cooperation in Taiwan as a SRAM/DRAM Design Manager. From 1998 to 2000, he was with Macronix International Cooperation in Taiwan as a Flash Design Manager. He is presently with eMemory Technology Incorporation, HsinChu, Taiwan. He is currently responsible for the design of standalone and embedded nonvolatile memory including Flash, MTP, and OTP products and IP.

Constant-Charge-Injection Programming: A Novel High-Speed Programming Method for Multilevel Flash Memories

Hideaki Kurata, Shunichi Saeki, Takashi Kobayashi, Yoshitaka Sasago, Tsuyoshi Arigane, Kazuo Otsuga, and Takayuki Kawahara, *Senior Member, IEEE*

***Abstract*—Constant-charge-injection programming (CCIP) has been proposed as a way to achieve high-speed multilevel programming in flash memories. In order to achieve high programming throughput in multilevel flash memory, programming method must provide: 1) high-speed cell-programming; 2) high programming efficiency; and 3) highly uniform programming characteristics. Conventional source-side channel-hot-electron injection (SSI) programming realizes both fast cell-programming and high programming efficiency, but the large cell-to-cell variation in programming speed with SSI is a problem. CCIP reduces the characteristic variation of SSI programming and satisfies all of the above requirements. By applying CCIP to 2-bit/cell AG-AND flash memory, the high programming throughput of 10.3 MB/s is obtained with 3% area penalty. This is 1.8 times faster than the throughput with conventional SSI programming.**

***Index Terms*—AG-AND, CCIP, flash memory, high-speed programming, multilevel cell, SSI.**

I. Introduction

THE increasing application of flash memory as the main storage medium of portable equipment such as digital still cameras and music players is creating requirements for greater storage capacities and faster programming. Storage capacities above 100 MB are required for the storage of high-resolution pictures in digital cameras, still or moving, and for CD-quality music recording in digital audio players. In addition, if we set a target of 10 s for downloading 100 MB of music data (data in MP3 audio format that plays for time equivalent to that of a single CD), the required programming throughput is 10 MB/s.

The multilevel cell (MLC) technique, in which two bits are stored in each physical memory cell [1], [2], is one of the most effective approaches for expanding storage capacity. When multilevel programming is used, however, two main factors slow down the programming throughput [3]–[6]. One is the large swing of $V_{\rm th}$, which extends the cell-programming time. The other is that careful adjustment takes time to narrow the mid-level $V_{\rm th}$ distributions by repeated programming and verification.

Manuscript received April 20, 2004; revised September 1, 2004.

H. Kurata, T. Kobayashi, Y. Sasago, T. Arigane, K. Otsuga, and T. Kawahara are with the Central Research Laboratory, Hitachi, Ltd., Kokubunji, Tokyo 185-8601, Japan (e-mail: h-kurata@crl.hitachi.co.jp).

S. Saeki is with Renesas Northern Lapan Semiconductor, Inc., Kodaira, Tokyo 187-8588, Japan.

Digital Object Identifier 10.1109/JSSC.2004.841019

The programming throughput (PT) of multilevel flash memories in general is expressed by

$$PT = \frac{N_{\rm bit}}{T_{\rm cell} + N_{\rm bit}/f_{\rm clock} + T_{\rm set} + T_{\rm vfy} \times N_{\rm vfy}} \quad (1)$$

where $T_{\rm cell}$ is the cell-programming time, $N_{\rm bit}$ is the number of cells being programmed simultaneously and $f_{\rm clock}$ is the clock frequency of the interface. $T_{\rm set}$ is the time overhead which is not related to verification, including the time taken to set up the internal programming voltages. $T_{\rm vfy}$ is the time overhead for each verification, and $N_{\rm vfy}$ is the number of internal programming and verification cycles for one programming operation. While three of the parameters in (1), $f_{\rm clock}$, $T_{\rm set}$ and $T_{\rm vfy}$, depend principally on the peripheral circuits, the other three parameters, $N_{\rm bit}$, $T_{\rm cell}$ and $N_{\rm vfy}$, are strongly dependent on the cell-programming method. To achieve high programming throughputs for multilevel flash memories, a large $N_{\rm bit}$, short $T_{\rm cell}$, and small $N_{\rm vfy}$ are indispensable.

Programming of a multilevel flash memory cell to the highest level requires a large $V_{\rm th}$ shift of 4 V, which is about 1.5 times as great as the shift required in a two-level flash memory. High-speed cell programming, that is, a short $T_{\rm cell}$, is thus essential.

We can program many cells at a time, if the current consumption of one memory cell during programming is small. Programming efficiency is the ratio of the injection current to the channel current (current drawn). If we are to further increase $N_{\rm bit}$, we need to raise the programming efficiency.

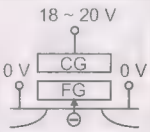
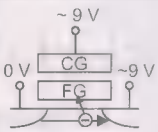
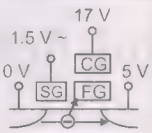
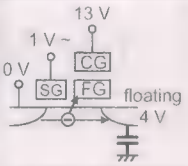
In response to a single external program command, the mid-level $V_{\rm th}$ distributions in a multilevel flash memory are sharpened by subjecting cells that fall outside the desired distributions to repeated cycles of internal programming and verification. A larger cell-to-cell variation in programming characteristics means a larger $N_{\rm vfy}$ and correspondingly poorer programming performance.

Thus, for a large $N_{\rm bit}$, short $T_{\rm cell}$, and small $N_{\rm vfy}$, the cell-programming method must provide: 1) high-speed cell-programming; 2) high programming efficiency; and 3) highly uniform programming characteristics. However, no conventional programming method satisfies all of the above requirements. Table I gives a comparison of programming methods.

A. Fowler–Nordheim (FN) Tunneling

FN tunneling is used in the programming of conventional AND- [7] and NAND-type [8] multilevel flash memories. The

TABLE I
COMPARISON OF PROGRAMMING METHODS

	FN tunneling	CHE injection	SSI	CCIP
Bias condition				
Cell speed	× 50 ~ 200 μs	✓ ~ 10 μs	✓ ~ 10 μs	✓ ~ 10 μs
Prog. efficiency (Prog. parallelism)	✓ ~ 1 (~ kB)	× ~ 10^{-6} (~ byte)	✓ ~ 10^{-3} (~ kB)	✓ ~ 10^{-3} (~ kB)
Distribution	× ~ 2.5 V	✓ ~ 1.5 V	× ~ 4 V	✓ ~ 1.5 V

advantage of this method is its high programming efficiency, which allows programming parallelism on the kilobyte scale and increases overall programming throughput. However, FN tunneling requires cell-programming times of 50 μs and longer, as well as strong electric fields during programming. In addition, the programming characteristics (threshold voltage distributions) are not uniform because they are highly sensitive to certain device parameters, such as the gate-coupling ratio [9], [10]. As is shown in Table I, the $V_{\rm th}$ distribution of memory cells programmed through FN tunneling with no internal reprogramming and verification is a large 2.5 V. Therefore, the long $T_{\rm cell}$ and large $N_{\rm vfy}$ limit the programming throughput.

B. CHE Injection

Channel-hot-electron (CHE) injection is realizable in simple stacked-gate devices and is thus widely used in the programming of NOR-type flash memories. This method achieves both high-speed cell-programming (10 μs) and high uniformity of programming, with a $V_{\rm th}$ distribution of only about 1.5 V [11]. The major drawback of CHE injection is its low programming efficiency ($\leq 10^{-6}$), which is due to the incompatibility between the optimal conditions for high hot-carrier generation and for electron collection on the floating gate. Therefore, CHE injection cannot achieve sufficient programming throughput for media-storage, because $N_{\rm bit}$ in (1) is only several bytes. On the other hand, NOR-type flash memories are used mainly for code storage, and in this application the low programming throughput of CHE injection does not affect the performance.

C. SSI

Source-side channel-hot-electron injection (SSI) [12]–[15] is the most suitable method in terms of both fast cell-programming ($\approx$ 10 μs) and good programming parallelism ($\approx$ kB). As the conditions for generating large numbers of hot carriers and strong injection can be made consistent, SSI programming achieves high programming efficiencies of more than 10^{-3}. However, the problem with SSI is the large variation in programming characteristics ($V_{\rm th}$ is distributed across more than 4 V). To achieve high-speed programming in multilevel flash memories, this variation must be reduced.

In this paper, we describe constant-charge-injection programming (CCIP), which realizes high-speed multilevel programming in flash memories. With CCIP, we achieve fast and precise control of $V_{\rm th}$ by suppressing the characteristics variation of SSI programming. By utilizing CCIP, we obtained a short $T_{\rm cell}$ of 10 μs, large $N_{\rm bit}$ of 8 kB, and small $V_{\rm th}$ variation of 1.5 V. Furthermore, applying CCIP to AG-AND multilevel flash memory achieved a programming throughput above 10 MB/s.

In Section II, we describe the mechanism of SSI and the problem with this method in terms of high-speed multilevel programming. Next, the concept of CCIP is presented in Section III. We then examine the application of CCIP to AG-AND flash memory in Section IV. The experimental results measured for a 32-Mb test chip are given in Section V. In Section VI, we discuss potential problems of leakage current. Section VII presents our estimation of performance for a 1-Gb AG-AND flash memory to which we apply CCIP. Finally, we conclude with a brief summary in Section VIII.

II. THE PROBLEM WITH SSI PROGRAMMING

SSI programming realizes high programming efficiency and fast cell-programming. The large cell-to-cell variation in programming speed with SSI is, however, a problem. In this section, we discuss the mechanism of SSI and the problem of variation in programming speed in terms of high-speed multilevel programming.

A. High Programming Efficiency of SSI Programming

SSI programming was developed as a way to obtain high programming efficiency. In the pioneering PACMOS (perpendicularly accelerating channel injection MOS) concept [12], a high potential at the floating gate is achieved by strong coupling with the drain. Since the potential of the floating gate can never be above that of the drain, conditions are not optimal for the collection of electrons on the floating gate.

The split triple-gate concept [13]–[15] was developed as a way to realize both the generation of large numbers of hot carriers and strong injection. Fig. 1 is a schematic diagram of the split triple-gate structure. An additional polysilicon select gate, such as a sidewall gate, is placed on the source side of the

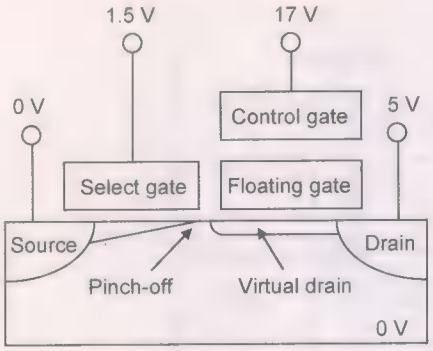


Fig. 1. Schematic view of split triple-gate flash memory programming.

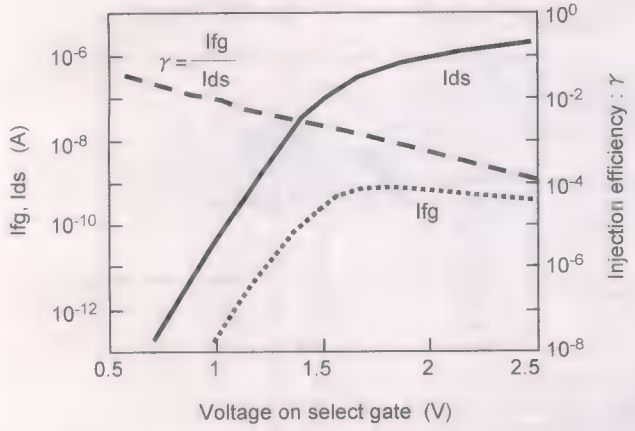


Fig. 2. Dependence of $I_{\rm fg}$, $I_{\rm ds}$, and injection efficiency on the voltage on the select gate.

floating gate. Typical internal operating voltages are 17 V for the control gate, 5 V for the drain, and 1.5 V for the select gate. This programming bias condition creates a virtual drain, which is an extension of the drain potential through the inversion layer beneath the floating gate. As a result, a pinch-off condition appears at the boundary between the select gate and the floating gate, which enhances the generation of hot electrons. Some of these hot electrons are injected into the floating gate by the vertical electric field at the pinch-off point.

The dependence of channel current ($I_{\rm ds}$) and injection current ($I_{\rm fg}$) on select gate bias is shown in Fig. 2. This was measured for an AG-AND flash memory unit [16], an extension of the split triple-gate structure. Further details are given in Section IV. Achieving 10 μs cell-programming requires a large injection current of more than 70 pA. $V_{\rm sft}$, the $V_{\rm th}$ shift of the memory cell due to a single programming pulse, i.e., a single internal programming operation, is given by

$$V_{\rm sft} = \frac{Q_g}{C_{\rm fg} \times R_c} = \frac{I_{\rm fg} \times T_{\rm cell}}{C_{\rm fg} \times R_c} \tag{2}$$

where Q_g is the total injection charge, $C_{\rm fg}$ is the total capacitance of the floating gate, and R_c is the coupling ratio of the control gate to the floating gate. In multilevel flash memory, a large $V_{\rm sft}$ of about 4 V is required. In this case, as $C_{\rm fg}$ is about 0.3 fF and R_c is about 0.6, a large $I_{\rm fg}$ of 70 pA is necessary to achieve a short $T_{\rm cell}$ of 10 μs.

On the other hand, in order to achieve more than kilobyte parallel programming, $I_{\rm ds}$ should be no more than 100 nA. This is because current supply from the internal voltage source is limited to about 10 mA, due to restrictions on chip area and current consumption.

As is shown in Fig. 2, both $I_{\rm fg}$ greater than 70 pA and $I_{\rm ds}$ less than 100 nA can be made consistent when the voltage of select gate is about 1.2 V. A high programming efficiency of more than 3×10^{-3} had thus been obtained; this is about three orders of magnitude better than the value for a conventional stacked-gate structure. Therefore, by utilizing SSI programming with a split triple-gate structure, both fast cell-programming and programming parallelism above the kilobyte scale are accomplished in combination with low power consumption.

B. Variation in Programming Speed

Here, we show the problem with SSI programming, i.e., the variation in programming speed. As is shown in Fig. 2, achieving fast cell-programming with low channel current requires that the select gate be operated in the subthreshold region. So, $I_{\rm ds}$ varies exponentially with linear variation in the $V_{\rm th}$ of MOS transistors formed under the select gate. This variation in $I_{\rm ds}$ leads to variation in programming speed. The charge injected into the floating gate (Q_g) is expressed as

$$Q_g = \int_0^t \gamma \times I_{\rm ds}\, dt \tag{3}$$

where γ is the programming efficiency. In (3), $I_{\rm ds}$ is almost constant during the programming pulse. If we define the average programming efficiency during the whole period of programming bias as γ_1, the expression for Q_g can be rewritten as

$$Q_g = \gamma_1 \times I_{\rm ds} \times t. \tag{4}$$

The $V_{\rm th}$ variation of select gate transistors is assumed to be ± 0.2 V in 130-nm manufacturing processes. Therefore, $I_{\rm ds}$ varies by more than two orders of magnitude, which produces a large variation of programming speed. This variation increases the number of internal programming and verification operations, $N_{\rm vfy}$, and degrades the programming performance. $N_{\rm vfy}$ is expressed as

$$N_{\rm vfy} \geq \frac{V_{\rm diff}}{\Delta V_{\rm th}} \tag{5}$$

where $V_{\rm diff}$ is the $V_{\rm th}$ difference between the fastest cell and slowest cell those are programmed without verification. $\Delta V_{\rm th}$ is the $V_{\rm th}$ distribution that is intended after verification, which is about 0.2 V. In multilevel flash memory, a sharp $V_{\rm th}$ distribution is formed by the repetition of both programming and verification. So a large variation of programming characteristics increases $N_{\rm vfy}$ and degrades programming performance. For example, when $V_{\rm diff}$ is 4 V, $N_{\rm vfy}$ is required to be a high 20 for every $V_{\rm th}$ level. To reduce $N_{\rm vfy}$, we have to decrease $V_{\rm diff}$. The target value for $V_{\rm diff}$ is less than 1.5 V, which will reduce $N_{\rm vfy}$ from 20 to 8 times. CCIP [17] has been developed as a method to suppress the variation of SSI programming.

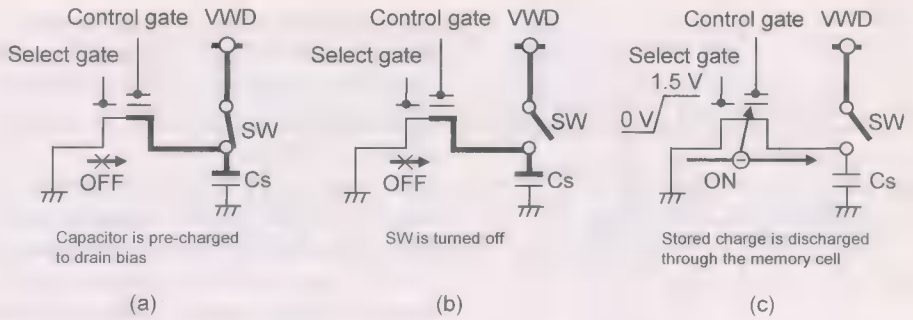


Fig. 3. Concept of CCIP. (a) Step 1. (b) Step 2. (c) Step 3.

III. Constant-Charge-Injection Programming

In conventional SSI programming, variation in $I_{\rm ds}$ leads to variation in programming characteristics. The essential point of CCIP is that the total amount of charge flowing through each memory cell in each programming operation is kept constant. This leads to the injection of constant charge into each floating gate. To obtain this constant flow of charge, each cell has to be equipped with a capacitor and switch.

The concept of CCIP is shown in Fig. 3. The capacitor (Cs) is attached between ground and the drain node of the memory cell. The switch (SW) connects the drain node with VWD, which is the internal power supply for drain bias, V_{wd}. CCIP is performed in three steps, with the aid of the capacitor and switch. In the first step, the switch is turned on and the capacitor is connected to VWD. The capacitor is then charged to V_{wd}, which is about 5 V. In the second step, which takes place when the voltage across the capacitor has reached V_{wd}, the switch is turned off. In the third step, the voltage on the select gate is raised to the programming bias. The charge stored in the capacitor is then discharged through the memory cell, generating hot electrons which are injected into the floating gate. The total charge injected into the floating gate (Q_g) is expressed as

$$Q_g = \int_0^{V_{wd}} \gamma \times Cs \, dV \tag{6}$$

If we define the average programming efficiency across the whole range of drain bias from V_{wd} to 0 V as γ_2, (6) may be rewritten as

$$Q_g = \gamma_2 \times Cs \times V_{wd} = \gamma_2 \times Q_o \tag{7}$$

where Q_o is the total charge stored in the capacitor. The dominant factor in variation of the capacitance of Cs is relatively small (less than $\pm 5\%$ of Cs), so Q_o can be regarded as almost constant. In addition, as is shown in Fig. 2, the variation in γ is about 0.2 of a decade under the condition that $V_{\rm th}$ variation of the select gate transistor is ± 0.2 V. Since γ is much less dependent on the select gate bias than $I_{\rm ds}$, we can obtain a near-constant Q_g. Therefore, CCIP realizes uniform programming by suppressing the variation of programming speed in SSI programming. In the next section, we discuss the application of CCIP to 130-nm AG-AND flash memory.

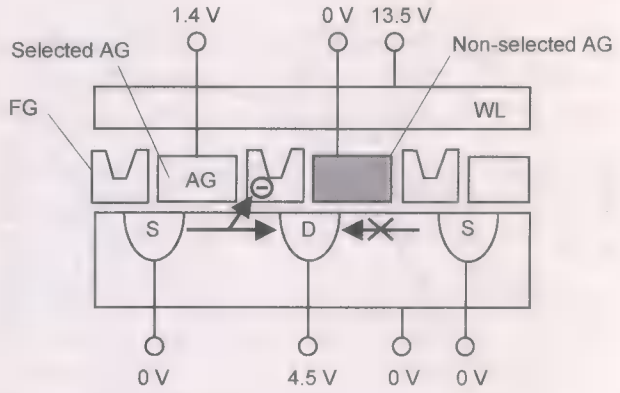


Fig. 4. Schematic view of AG-AND flash memory programming.

IV. Applying CCIP to AG-AND Flash Memory

A. AG-AND Flash Memory

Schematic diagrams of the memory cell and array architecture of AG-AND flash memory are given in Figs. 4 and 5. The assist gate (AG) is equivalent to the select gate of Fig. 1. The memory array is a virtual-ground structure and 256 memory cells are connected in parallel to the local bit-lines, each of which is a diffusion layer.

Selection transistors control connection of the local bit-lines to the global bit-lines. One set of assist-gates ($\rm AG_e$) acts as the program gates for the selected memory cells (A and C in Fig. 5) while the other set, $\rm AG_o$, acts as the field-isolation gates for the nonselected transistors (B in Fig. 5). The AG set to which the respective AG lines belong alternates across the structure, and the lines are joined up just beyond the ends of the local bit-lines. To reduce the data-line pitch, the floating gates were embedded in the spaces between the AGs by a self-aligned process. The floating gates have a three-dimensional shape, which enhances the coupling ratio with the word-lines. The unit cell area is 0.104 μm^2, the data-line pitch is 0.4 μm, and the word-line pitch is 0.26 μm.

Bias conditions for programming, erasure, and reading are listed in Table II. For erasure, a negative bias is applied to the selected word-line. Under this condition, electrons flow from the floating gates to the substrate by FN tunneling.

The memory cell is programmed by source-side channel-hot-electron injection for high programming efficiency. The internal operating voltages are 13.5 V for the selected word-line (WL),

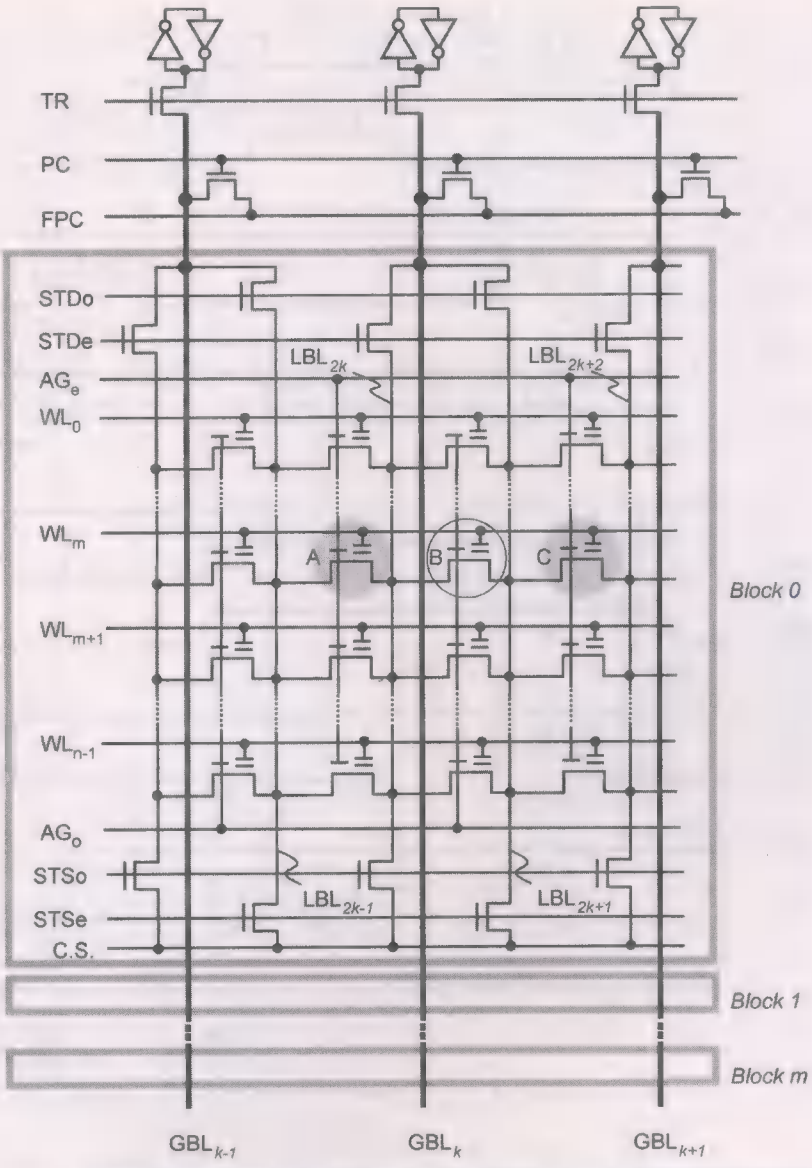


Fig. 5. Array architecture of AG-AND.

TABLE II
BIAS CONDITIONS FOR PROGRAMMING, ERASING, AND READING

	WL$_m$	WL$_{m+1}$	AG$_o$	AG$_e$	LBL$_{2k-1}$	LBL$_{2k}$	LBL$_{2k+1}$	LBL$_{2k+2}$
Prog.	13.5 V	0 V	1.6 V	0 V	0 V	4.5 V	0 V	4.5 V
Erase	-17 V	0 V	0 V	0 V	0 V	0 V	0 V	0 V
Read	Vread	0 V	3 V	0 V	0 V	1 V	0 V	1 V

4.5 V for the drain, and 1.4 V for the selected AG. During programming of cell A in Fig. 4, the AG of cell B (AGo) is kept at 0 V to suppress channel formation.

B. Operation of CCIP

As was described in Section III, realizing CCIP requires the addition of a capacitor and a switch to each of the selected memory cells, and this leads to a large increase in chip area.

To achieve CCIP operation for an AG-AND flash memory with no penalty in terms of chip area, we use the stray capacitance of the diffusion local bit-line as the capacitor and the selection transistor as the switch. The stray capacitance of the local bit-line is 40 fF, which is largely composed of the capacitance of the p-n junction. The $V_{\rm th}$ shift for a memory cell in response to a single programming pulse ($V_{\rm sft}$) is expressed as

$$V_{\rm sft} = \frac{Q_g}{C_{\rm fg} \times R_c} = \frac{\gamma_2 \times C_s \times V_{wd}}{C_{\rm fg} \times R_c} \tag{8}$$

where $C_{\rm fg}$ is the total capacitance of the floating gate and R_c is the coupling ratio of the control gate to the floating gate. As $C_{\rm fg}$ is about 0.3 fF and R_c is about 0.6, $V_{\rm sft}$, the change in threshold voltage with a single programming pulse is about 3.0 V.

The timing diagram of CCIP is shown in Fig. 6, which applies to programming of the hatched cells in Fig. 5. In the first

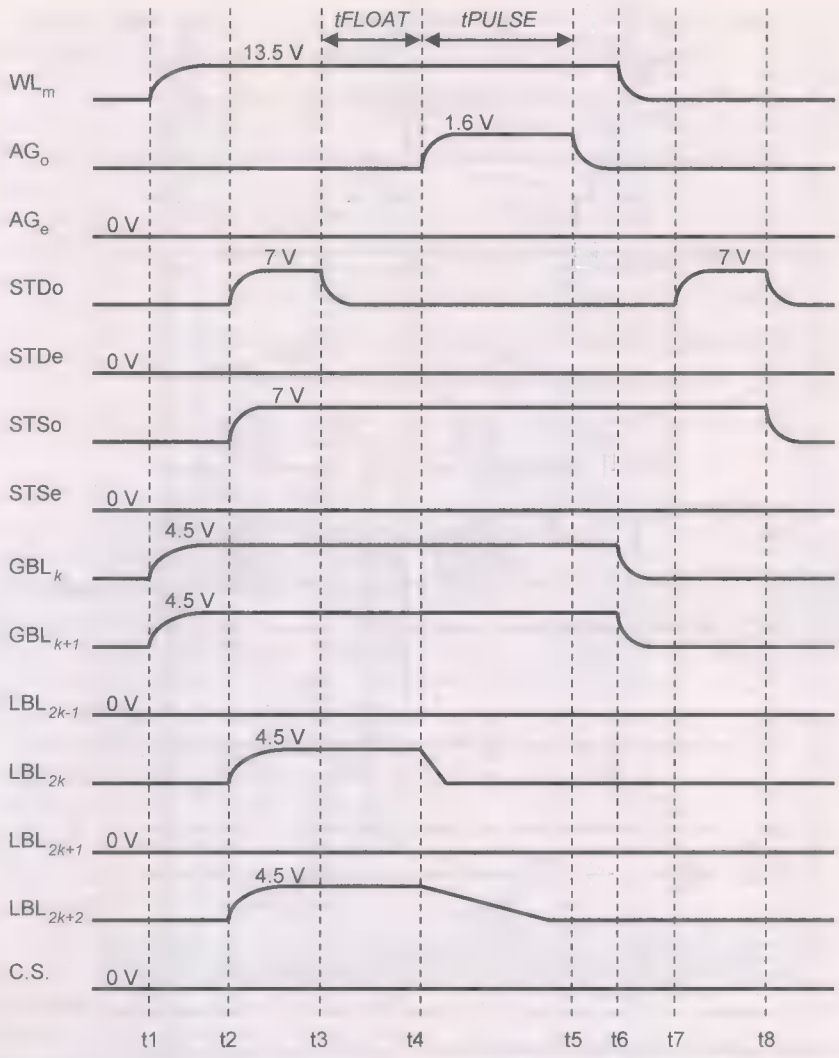


Fig. 6. Timing diagram of CCIP.

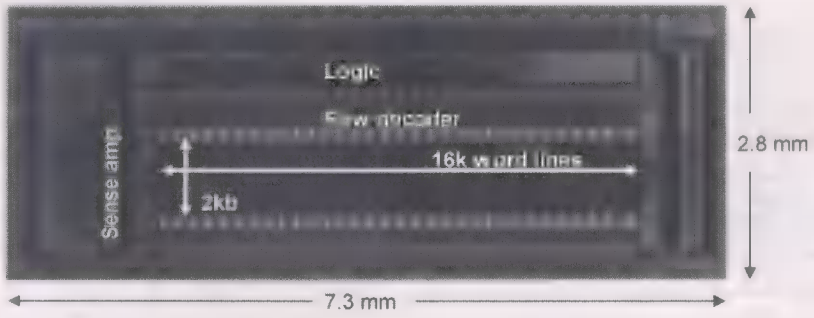


Fig. 7. Microphotograph of 32-Mb AG-AND flash test chip.

step (at *t2*), the gate signal of the relevant selection transistors (STDo) becomes high and the local bit-lines (LBL_{2k} and LBL_{2k+2}) are charged to 4.5 V. After charging of the local bit-lines is completed (at *t3*), the selection transistors are turned off. Local bit-lines LBL_{2k} and LBL_{2k+2} are then floating. Finally, when AGo becomes high at *t4*, the stored charge in LBL_{2k} and LBL_{2k+2} is discharged through cell A and cell C. The pulse width (*tPULSE*) must be long enough for the slowest cell to discharge all of its stored charge.

V. Experimental Results

A 32-Mb AG-AND test chip was fabricated in 0.13 μm CMOS technology and is shown in Fig. 7. Key device characteristics and parameters are summarized in Table III. A triple-well CMOS process on a p-type substrate was used. The tunnel oxide of the memory cells is 9 nm thick and the gate-oxide layers of the high- and low-voltage peripheral transistors are 25 nm and 9 nm thick, respectively. The word-line

TABLE III
DEVICE FEATURES

Process	: 0.13 um p-sub CMOS triple-well 2 poly-Si, 1 W, 2 Al
Gate oxide	: 25 nm (H.V.) , 9 nm (L.V.)
Tunnel oxide	: 9 nm
Interpoly dielectric	: 14 nm
Cell size	: 0.052 μm^2 / bit

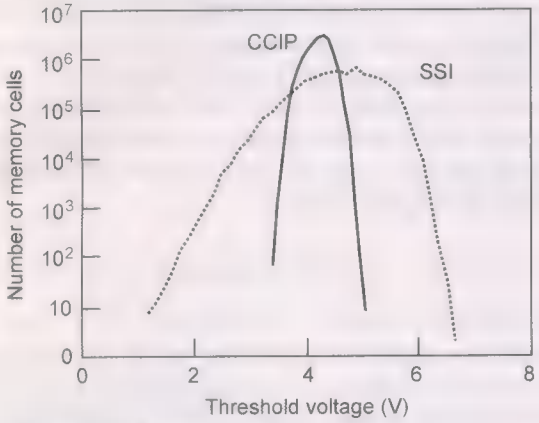


Fig. 8. Distributions of programming characteristics.

pitch is 0.26 μm and the bit-line pitch is 0.4 μm. The bit area of the cell is 3.1 F^2, for a value of 0.052 μm^2 with the 0.13 μm process. Key points from the results of measurement of this test chip are given below.

Comparable results on V_{th} distribution for conventional SSI programming and CCIP are given in Fig. 8. The number of measured cells is equivalent to 4 Mb. As is shown in Fig. 8, the V_{th} distribution with conventional SSI programming spans a broad 5.5 V (V_{th}: 1.0–6.5 V). By utilizing CCIP, however, we dramatically narrow the V_{th} distribution to span less than 1.5 V (V_{th}: over 3.5 to 5.0 V).

Figs. 9 and 10 show the programming characteristics. The *X*-axis indicates the number of internal programming pulses. The length of each pulse is 1 μs. These figures indicate that controlling the word-line voltage (V_{ww}) and the drain voltage (V_{wd}) are both effective as ways to optimize the programming speed.

VI. CHARGE LEAKAGE

This section covers potential problems of leakage that accompany the proposed scheme. When the local bit-line is precharged to the programming voltage, we see two kinds of charge leakage from the floating drain node. The first is a p-n junction leakage and the second is a gate-induced-drain leakage. Since charge leakage reduces Q_o in (4), it also lowers the programming speed.

A. Junction Leakage

Since the storage capacitor is a p-n junction, it has p-n junction leakage. This leakage is determined by the breakdown voltage of the p-n junction (BVj). Fig. 11 shows how BVj affects the dependence of the programming characteristics (drop in V_{th}) on *tFLOAT*, which is the period over which the local bit-line floats, as shown in Fig. 6. With increasing *tFLOAT*, the leakage current increasingly lowers V_{th}, so that more rounds of reprogramming and verification are required, leading to lower programming speeds. The results indicate that increasing BVj is highly effective as a way of suppressing the programming degradation.

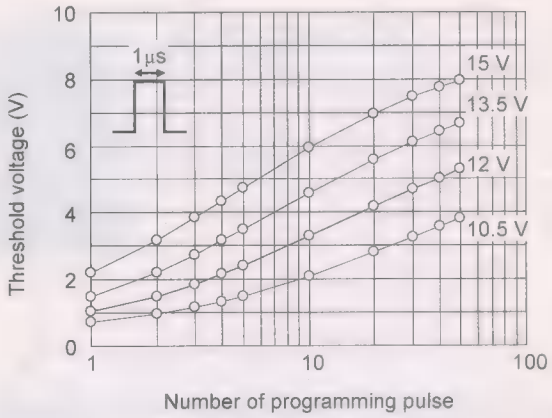


Fig. 9. Change of programming characteristics with word-line voltage.

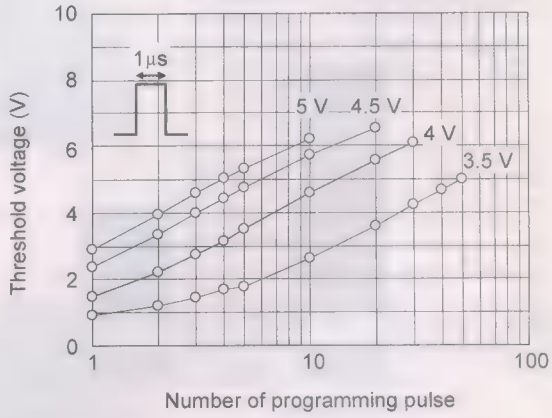


Fig. 10. Change of programming characteristics with drain voltage.

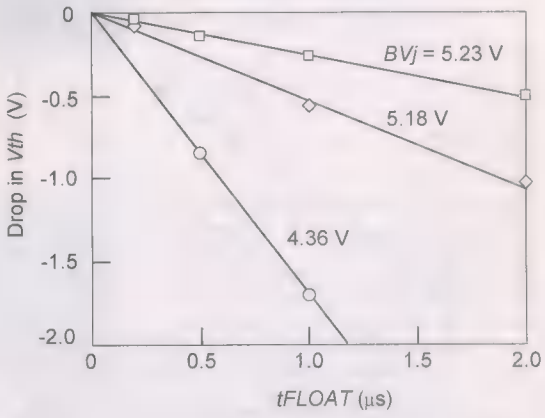


Fig. 11. Effect of p-n junction leakage.

B. Gate-Induced-Drain Leakage (GIDL)

GIDL is caused by band-to-band tunneling in the gate-overlap region of the drain. High values for GIDL are obtained by a high

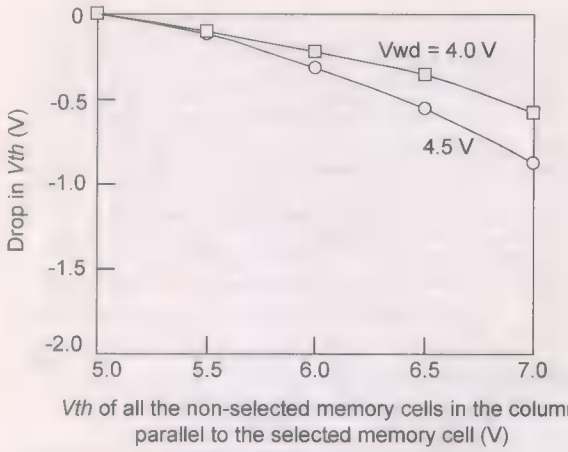


Fig. 12. Effect of gate-induced drain leakage.

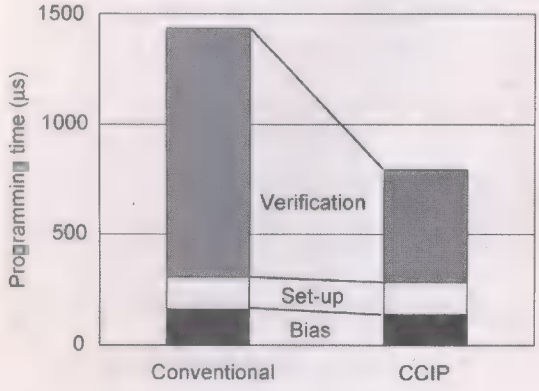


Fig. 13. Comparison of programming times.

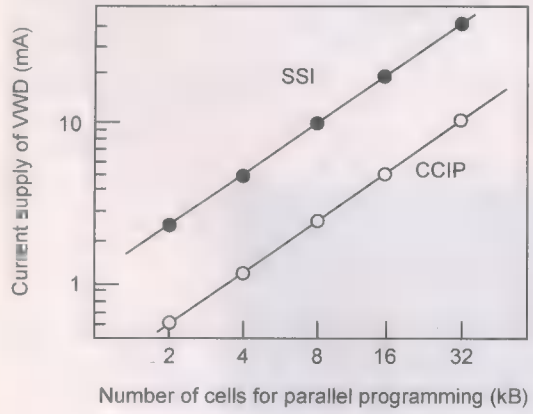


Fig. 14. Current supply of VWD.

$V_{\rm th}$ for all the nonselected memory cells in the columns parallel to the selected memory cells (Fig. 12). Therefore, suppressing the multilevel $V_{\rm th}$ window is effective to suppress the programming degradation.

VII. Performance of 1-Gb AG-AND Flash

In this section, we present estimates of the programming performance of 1-Gb multilevel AG-AND flash memory units with SSI and CCIP programming.

Fig. 13 compares programming times. When conventional SSI programming is applied, the deviation in programming characteristic increases the number of internal programming and verification cycles ($N_{\rm vfy}$) so that this process alone requires almost 1 ms. CCIP decreases $N_{\rm vfy}$ by lowering the variation of threshold voltage relative to that seen in SSI programming. The time overhead of verification is reduced to 45% of the value for SSI. Given 8-kB-parallel programming, a programming throughput of 10.3 MB/s (i.e., 1.8 times faster than with SSI) is achieved.

In addition to high-speed programming and uniformity of programming characteristics, CCIP has the advantage of a lower current requirement for programming than SSI (Fig. 14). In SSI programming, channel current flows through the memory cells during the entire programming period. So, in the case of 8-kB parallel programming, a current source providing more than 10 mA is required for VWD. However, the internal voltage source only has to pre-charge the capacitor from the bit-line. That is, the bit-line voltage only has to drive 35% of the current required with SSI programming.

VIII. Conclusion

Constant-charge-injection programming (CCIP) has been proposed as a method for the high-speed programming of multilevel flash memories. As a replacement for conventional FN, CCIP based on SSI is a key technology for high-speed multilevel programming. By utilizing CCIP, we obtained high-speed cell programming of 10 μs, high programming efficiency of more than 3×10^{-3}, and high uniformity of programming, with a $V_{\rm th}$ distribution of 1.5 V. AG-AND flash memory with the proposed scheme enables 10.3 MB/s multilevel programming.

Acknowledgment

The authors thank K. Furusawa, O. Tsuchiya, A. Nozoe, K. Izawa, M. Kanamitsu, K. Yoshida, Y. Sakamoto, J. Kishimoto, Y. Takase, M. Sakai, T. Fujimoto, T. Bando, H. Kume, and K. Kimura for their encouragement and suggestions.

References

[1] A. Nozoe *et al.*, "A 256 Mb multilevel flash memory with 2 MB/s program rate for mass storage application," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 1999, pp. 110–111.

[2] T. Cho *et al.*, "A dual-mode NAND flash memory: 1-Gb multilevel and high-performance 512-Mb single-level modes," *IEEE J. Solid-State Circuits*, vol. 36, no. 11, pp. 1700–1706, Nov. 2001.

[3] H. Nobukata *et al.*, "A 144-Mb, eight-level NAND flash memory with optimized pulsewidth programming," *IEEE J. Solid-State Circuits*, vol. 35, no. 5, pp. 682–690, May 2000.

[4] H. Kurata *et al.*, "A selective verify scheme for achieving a 5-MB/s program rate in 3-bit/cell flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2000, pp. 166–167.

[5] T.-S. Jung *et al.*, "A 117-mm^2 3.3-V-only 128-Mb multilevel NAND flash memory for mass storage applications," *IEEE J. Solid-State Circuits*, vol. 31, no. 11, pp. 1575–1583, Nov. 1996.

[6] K. Takeuchi *et al.*, "A multipage cell architecture for high-speed programming multilevel NAND flash memories," *IEEE J. Solid-State Circuits*, vol. 33, no. 8, pp. 1228–1238, Aug. 1998.

[7] H. Kume *et al.*, "A 1.28 μm^2 contactless memory cell technology for 3 V-only 64 Mbt EEPROM," in *IEDM Tech. Dig.*, 1992, pp. 991–993.

[8] F. Masuoka *et al.*, "New ultra high density EPROM and flash EEPROM with NAND structure cell," in *IEDM Tech. Dig.*, 1987, p. 552.

[9] D.-C. Kim *et al.*, "A 2 Gb NAND flash memory with 0.044 μm^2 cell size using 90 nm flash technology," in *IEDM Tech. Dig.*, 2002, pp. 919–922.
[10] Y.-S. Yim *et al.*, "70 nm NAND flash technology with 0.025 μm^2 cell size for 4 Gb flash memory," in *IEDM Tech. Dig.*, 2003, p. 819.
[11] V. N. Kynett *et al.*, "A 90-ns one-million erase/program cycle 1-Mb flash memory," *IEEE J. Solid-State Circuits*, vol. 24, no. 11, pp. 1259–1264, Nov. 1989.
[12] M. Kamiya *et al.*, "EPROM cell with high gate injection efficiency," in *IEDM Tech. Dig.*, 1982, p. 741.
[13] A. T. Wu *et al.*, "A novel high-speed, 5-volt programming EPROM structure with source-side injection," in *IEDM Tech. Dig.*, 1986, p. 584.
[14] N. Naruke *et al.*, "A new flash-erase EEPROM cell with a sidewall select-gate on its source side," in *IEDM Tech. Dig.*, 1989, p. 603.
[15] J. V. Houdt *et al.*, "Analysis of the enhanced hot-electron injection in split-gate transistor useful for EEPROM," *IEEE Trans. Electron Devices*, vol. 39, no. 5, pp. 1150–1156, May 1992.
[16] T. Kobayashi *et al.*, "A giga-scale Assist-Gate (AG)-AND-type flash memory cell with 20-MB/s programming throughput for content-downloading application," in *IEDM Tech. Dig.*, 2001, p. 29.
[17] H. Kurata *et al.*, "Constant-charge-injection programming for 10-MB/s multilevel AG-AND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2002, pp. 302–303.
[18] Y. Sasago *et al.*, "10-MB/s multi-level programming of Gb-scale flash memory enabled by new AG-AND cell technology," in *IEDM Tech. Dig.*, 2002, p. 952.
[19] K. Yoshida *et al.*, "A 1 Gb multilevel AG-AND-type flash memory with 10 MB/s programming throughput for mass storage application," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2003, p. 288.

Hideaki Kurata was born in Osaka, Japan, on May 1, 1970. He received the B.S. and M.S. degrees in electronic engineering from Kyoto University, Kyoto, Japan, in 1993 and 1995, respectively.

Since he joined the Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, in 1995, he has been engaged in the research and development of flash memories.

Shunichi Saeki was born in Tokyo, Japan, on May 18, 1963. He received the B.S. degree in electrical engineering from Tokai University, Tokyo, Japan, in 1986.

In 1986, he joined Hitachi Device Engineering Co., Ltd., Chiba, Japan. From 1986 to 2003, he engaged in the research and development of flash memory circuit. In 2003, He joined Renesas Northern Japan Semiconductor, Inc, Tokyo, Japan, and has been engaged in the development of LCD devices.

Mr. Saeki is a member of the Institute of Electronics, Information, and Communication Engineers of Japan.

Takashi Kobayashi was born in Nagano, Japan, on December 10, 1961. He received the B.S. and M.S. degrees in metal processing from Tohoku University, Sendai, Japan, in 1984 and 1986, respectively.

He joined the Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, in 1986, where he was engaged in research on dielectric and insulating films. Since 1994, he has been engaged in the research and development of high-density flash memories.

Mr. Kobayashi is a member of the Japan Society of Applied Physics, and a member of the programming committees of the IEEE International Electron Devices Meeting (IEDM) and the Solid-State Devices Meeting (SSDM).

Yoshitaka Sasago was born in Chiba, Japan, on February 20, 1969. He received the B.E., M.E., and D.E. degrees in 1992, 1994, and 1997, respectively, from the University of Tokyo, Tokyo, Japan.

In 1997, he joined Hitachi Ltd., Tokyo. Since 1999, he has been engaged in the research and development of flash memory.

Tsuyoshi Arigane was born in Ibaraki, Japan, on July 30, 1971. He received the B.E., M.E., and D.E. degrees in 1995, 1997, and 2000, respectively, from the University of Tohoku, Miyagi, Japan.

In 2000, he joined Hitachi Ltd., Tokyo, Japan, where he has been engaged in the research and development of flash memory.

Kazuo Otsuga was born in Aichi, Japan, on January 1, 1978. He received the B.S. and M.S. degrees in physics from Osaka University, Japan, in 2000 and 2002, respectively.

He joined the Central Research Laboratory, Hitachi Ltd, Tokyo, in 2002. Since then, he has been engaged in the research and development of high-density flash memory.

Takayuki Kawahara (M'91–SM'98) received the B.S. and M.S. degrees in physics in 1983 and 1985 and the Ph.D. degree in electronics in 1993 from Kyusyu University, Fukuoka, Japan.

In 1985, he joined Central Research Laboratory, Hitachi Ltd., Tokyo. Since then, he has made fundamental contributions in many areas in the field of low-power high-speed memories. In the field of DRAM circuits, from 1985 to 1993, his major contributions concerned low-power, low-voltage circuits including subthreshold-current reduction by gate-source self-reverse biasing technique and an over-drive sense-amplifier scheme coupled with direct sensing. He also pioneered the charge-recycling scheme, which concept is now a widely applied to various circuits. In the field of flash memory, from 1994 to 1998, he and his team developed a bit-line clamped sensing scheme for fast sensing, a high-voltage generator scheme under a low-voltage supply, and a pioneering high-speed programming method. In addition, he engaged in the ultra-low-power system LSI project in the laboratory from 1999 to 2002. Currently, he is leading the research groups of SRAM, DRAM, and Nonvolatile memory. He was a visiting researcher at Electronics Laboratory (LEG), Swiss Federal Institute of Technology Lausanne (EPFL), from 1997 to 1998.

Dr. Kawahara was a guest editor of the Memory section of the special issue of the IEEE JOURNAL OF SOLID-STATE CIRCUITS, November 2002. He has been a member of the ISSCC program committee since 2000 (and the executive committee since 2004), and a program committee member of the Symposium on VLSI Circuits since 2003.

A 1-GHz Signal Bandwidth 6-bit CMOS ADC With Power-Efficient Averaging

Xicheng Jiang and Mau-Chung Frank Chang, *Fellow, IEEE*

***Abstract*—A 2-GS/s 6-bit ADC with time-interleaving is demonstrated in 0.18-μm one-poly six-metal CMOS. A triple-cross connection method is devised to improve the offset averaging efficiency. Circuit techniques, enabling a state-of-the-art figure-of-merit of 3.5 pJ per conversion step, are discussed. The peak DNL and INL are measured as 0.32 LSB and 0.5 LSB, respectively. The SNDR and SFDR have achieved 36 and 48 dB, respectively, with 4 MHz input signal. Near Nyquist input frequencies, the SNDR and SFDR maintain above 30 and 35.5 dB, respectively, up to 941 MHz. The complete ADC, including front-end track-and-hold amplifiers and clock buffers, consumes 310 mW from a 1.8-V supply while operating at 2-GHz conversion rate. The prototype ADC occupies an active chip area of 0.5 mm^2.**

***Index Terms*—Analog-to-digital converter (ADC), averaging, CMOS, interleaving, track/hold, triple-cross connection.**

I. Introduction

HIGH-SPEED ADCs are an integral part of high-performance systems such as disk drive read channels, fiber optical receiver front-end and data communication links using multilevel signaling (e.g., PAM and QAM). The main issues in the design of such ADCs include static and dynamic offset reduction, low supply-voltage operation, gain and speed optimization. Design tradeoffs between power, speed, and chip area further tighten the design requirements. It is also of particular importance that such ADCs be implemented in a standard CMOS process for easy integration with larger signal processing circuits.

This paper presents the design of a 6-bit 2-GS/s ADC implemented in a 0.18-μm CMOS technology. The ADC performance in a standard CMOS process is constrained by the threshold mismatch of the CMOS devices. The offset averaging method proposed in [1] is a powerful technique to alleviate its impact in preamplifier or comparator arrays [1]–[4]. Nonetheless, it still requires further modifications to correct for optimum averaging effects at the array boundaries. This work introduces a triple-cross connection method to improve the averaging efficiency. Combining such a technique with time-interleaving and open-loop front-end track-and-hold amplifiers (THAs), the converter achieves a figure-of-merit of 3.5 pJ per conversion step. Section II introduces the triple-cross connection method. Section III describes the ADC architecture and the THA circuit. Section IV presents experimental results obtained from the prototype ADC.

Manuscript received January 30, 2004; revised June 16, 2004. This work was supported by the Defense Advanced Research Projects Agency (DARPA) and MICRO.

X. Jiang was with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA. He is now with Broadcom Corporation, Irvine, CA 92618 USA (e-mail: xicheng@icsl.ucla.edu).

M.-C. F. Chang is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA.

Digital Object Identifier 10.1109/JSSC.2004.841033

II. Triple-Cross Connection Method

A. Boundary Issues

Averaging acts like a spatial filter that can reduce the offset of the preamplifier. Since it smoothes out the faster fluctuation more than the slower one, the differential nonlinearity (DNL) usually gets more improvement than the integral nonlinearity (INL). One way to implement averaging is by inserting ladder resistors between outputs of adjacent amplifiers [1]. The averaging technique, however, causes problems at the averaging network boundaries. In general, there are two issues with traditional averaging networks at the boundaries. First, zero-crossings drift from input reference voltage levels due to the asymmetrical nature of the boundary. At the edge, the zero-crossings shift inward due to the lack of amplifiers on the other side. This drift causes systematic nonlinearity errors. Second, the number of random components contributing to the averaging is diminished at the boundary. This counteracts the resulting DNL/INL improvement through the averaging. In other words, the standard deviation of the input referred offset at the boundary is larger than at the center. Comparing with the amplifier array center, the input linear range of the preamplifier at network edge covers about a half the number of preamplifiers that can contribute to averaging. State-of-the-art designs use either dummy amplifiers to preserve the characteristics of an infinitely long amplifier array [2], [3], or resort to the extra boundary termination circuits to suppress the zero-crossing shifts at the edges [4]. The dummy method can be made more effective when more dummies are used. For instance, 18 dummies are required for an averaging window that covers 18 amplifiers [3]. However, this makes the averaging method rather inefficient, since only a part of the amplifier array and the reference range are usable. The edge termination method consumes less power and area. However, it only restores the systematic errors when the averaging window is narrow and the boundary issue is less severe. Furthermore, these methods need significant amount of extra reference range, which represents a serious challenge for low-supply applications.

B. Triple-Cross Connection

To solve the boundary problem, the first step is to make sure the averaging resistor network is properly terminated. One way to achieve this goal, as suggested in most folding ADCs with resistive interpolation, is to cross-connect outputs at the network boundaries. This preserves the translational symmetry of the impulse response [5] of the resistor network, but the primary issues such as zero-crossing shifts and noneven averaging remain. The clipped outputs at the other boundary provide a strong force pulling the zero-crossings outward, far away from their ideal positions. These clipped amplifiers will not contribute threshold mismatch components [6] from their input differential pairs to

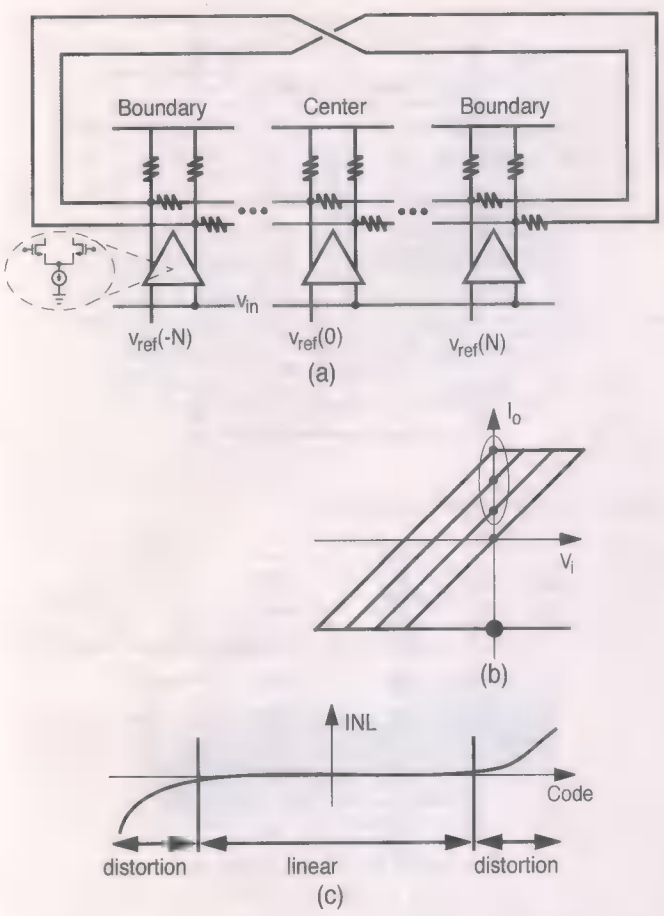


Fig. 1. (a) Preamplifier array with one cross connection at the boundary. (b) Preamplifiers from one side contribute to averaging at the edge. (c) INL profile.

averaging. Only a part of the array has the required linearity and both edges exhibit significant distortion, as shown in Fig. 1. Let us start with adding enough dummies at both boundaries. The over range references can be eliminated after observing the symmetry property of differential circuits. By cross-connecting outputs of the dummies to that of regular amplifiers, the dummies can be connected oppositely to existing reference points instead of over range references, as illustrated in Fig. 2. By doing so, we have achieved the following goals: 1) the extra references are eliminated; 2) the zero-crossing shifts are corrected due to symmetry being maintained; and 3) the input linear range at the boundary covers an equal number of amplifiers at the edge and at the center, which means the random offsets are averaged in the same scale from the array center to the boundary. However, the negative transconductances from the dummies reduce the effective transconductance at the boundary. Also, when the averaging window is wide, a significant number of dummies are required. This method can be further improved by designing an interface amplifier instead of using the regular preamplifiers as dummies. Like the regular preamplifier, the interface amplifier consists of an input differential pair, a current source, and resistive loads. The differential input devices are carefully sized such that the input linear region of the interface amplifier overlaps that of the adjacent regular preamplifier. The interface amplifier in Fig. 3 has a similar effect to the lumped ef-

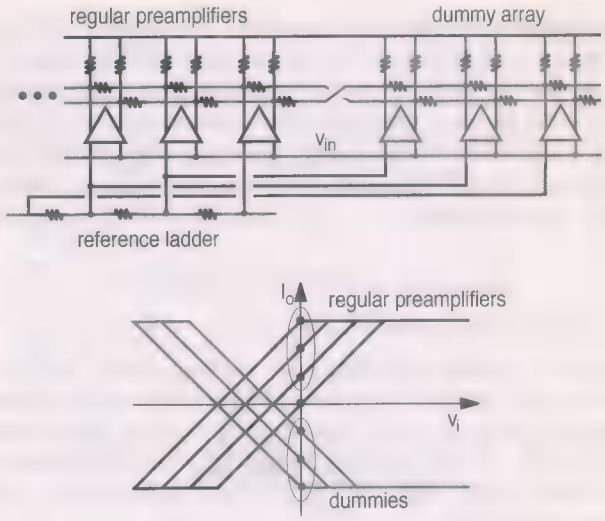


Fig. 2. Dummies without over range references.

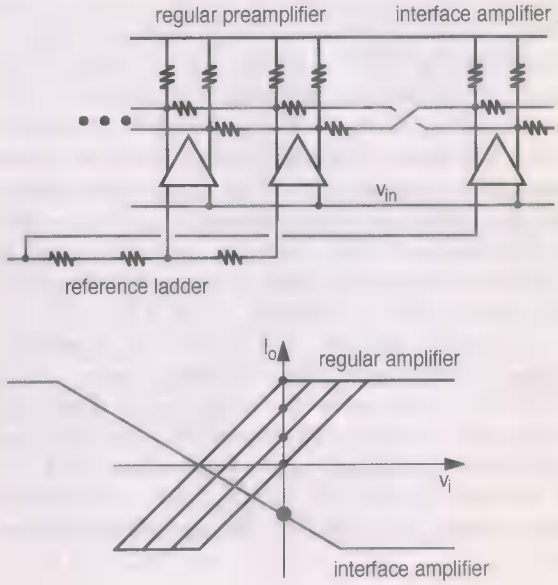


Fig. 3. Interface amplifier equivalent to the dummy array.

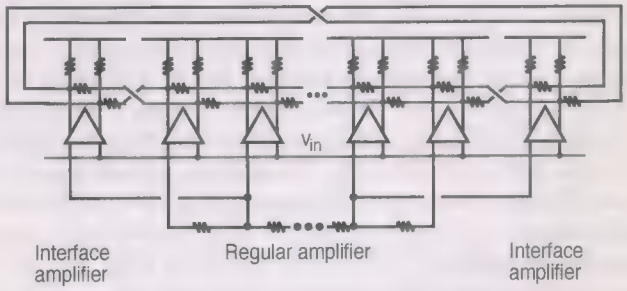


Fig. 4. Triple-cross connection scheme.

fect of the dummy array with respect to averaging. To minimize zero-crossing shifts and the negative transconductance, the reference point used for the interface amplifier is three steps away from the end of reference ladder. There is one interface amplifier at each network boundary. Fig. 4 shows the scheme of the triple-cross connection method. The two crossings at the bound-

aries minimize the zero-crossing shifts and the third crossing is for proper termination of the resistor network. Simulations indicate that the peak INL of the amplifier array can be reduced to 0.5 LSB by using the triple-cross connection method and interface amplifiers (down from 4.5 LSB with abrupt termination and from 7.2 LSB with only one cross-connection at the averaging network edges).

III. ADC Architecture and THA

A. Proposed ADC Architecture

In order to achieve the required data throughputs, time interleaving [7] is needed. It is used to relax the bandwidth requirements of individual ADC blocks (except for the THA, which still requires the full tracking bandwidth). This leads to higher ADC data throughput at a lower clock rate with reduced overall power consumption.

An open-loop THA with replica biasing is implemented to ensure the desired dynamic performance with broadband input signals. Interpolation is implemented at the comparator stage to save hardware and power consumption in preamplifiers. A current-mode logic (CML) comparator latch is used to lower the dynamic offset of the combined stage. The latch output swing is limited to 0.6 V (rather than 1.8 V rail-to-rail) to speed up the regeneration and reduce the dynamic offset. The gain required for suppressing the dynamic offset of the comparator is distributed among preamplifier stages to maximize the overall circuit bandwidth. The front-end THA decreases the bubble error probability arising from the clock skew. However, high-speed glitches remain a main source for bubble errors. The signal-to-noise ratio (SNR) drops and the output waveform is severely distorted due to these performance limiting glitches. A 3-input NAND following the comparators is used as the power-efficient error-reduction circuitry. A ROM-based encoder maps the thermometer code to the binary code. The detailed block diagram of the time-interleaving ADC with averaging and interpolation is shown in Fig. 5. The analog signal paths use fully differential circuits.

B. Track/Hold Amplifier

At gigahertz sampling frequencies, the THA [8] is critical for achieving good dynamic performance over broadband input signals. Fig. 6 shows an open-loop THA with replica-based "well-biasing." Source followers in the THA utilize sufficiently large PMOS devices to drive subsequent preamplifiers. The output of a small replica source follower is used to bias the well of the main source follower. This has linearity advantages over a source follower with a well-to-source connection, without the disadvantage of having that output drive the nonlinear well-substrate capacitance. The replica consumes only 5% of the power of the main source follower. The low input common-mode voltage reduces the on-resistance of the NMOS- switches and increases the input tracking bandwidth (−1 dB) to about 6.4 GHz. The dummy switches reduce the charge injection and the voltage glitch, thus reducing the dynamic offset.

IV. Experimental Results

Fabricated in a 0.18-μm one-poly six-metal (1P6M) CMOS technology, the chip microphotograph is shown in Fig. 7. The

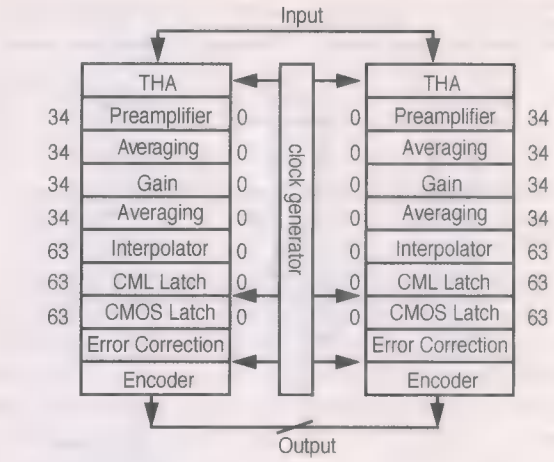


Fig. 5. 2 GS/s 6-bit ADC architecture.

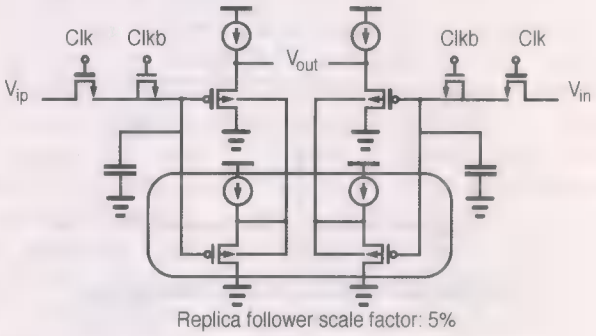


Fig. 6. Broadband THA with replica well-biasing.

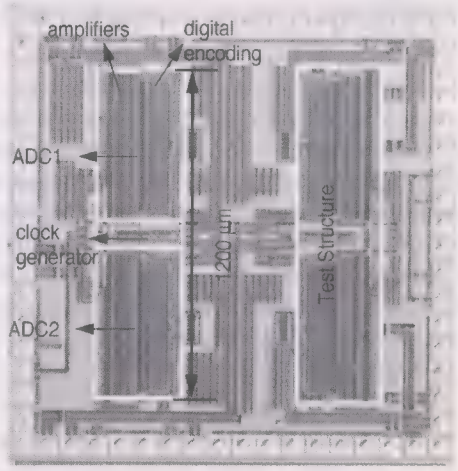


Fig. 7. Microphotograph of the fabricated 6-bit ADC.

right side contains a test structure and the left side contains the 2-GS/s 6-bit ADC. Two sub-ADCs are laid at the top and the bottom, respectively, with the clock generator and the buffer amplifier sitting at the center. For each of the sub-ADCs, from the left to the right, are amplifiers and digital encoders. The prototype ADC occupies an active chip area of 0.5 mm^2. A decoupling capacitor of about 1 nF is used to fill the empty space on the die. For easy testing, the ADC chip is mounted on a printed circuit board (PCB) with direct die-to-board wire bonding. There is no decimation at the ADC outputs. For dynamic analyses, the outputs from the two ADCs are combined

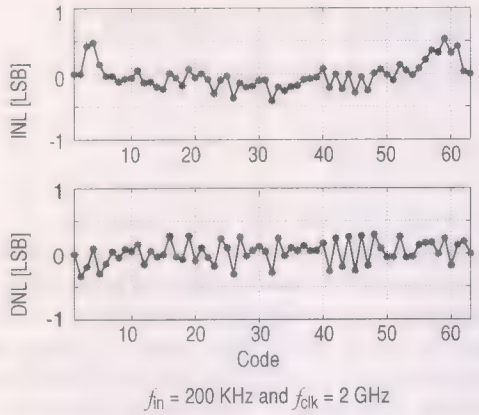


Fig. 8. Measured INL and DNL.

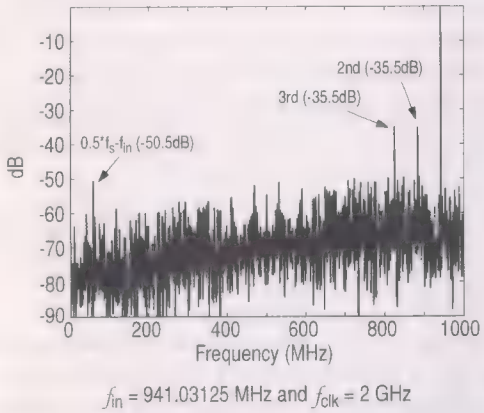


Fig. 9. Measured frequency spectrum.

to deliver the full 2-Gword/s rate. However, for static analyses, outputs from the two ADCs are separated to avoid numerical averaging. Fig. 8 shows the measured INL and DNL profiles. They are extracted from the histogram [9] of the 64 K ADC outputs in response to a 200-kHz sinusoidal input signal at the sampling rate of 2 GHz. The peak INL and DNL are recorded as 0.5 LSB and 0.32 LSB, respectively. This plot shows the systematic nonlinearity being corrected. When the input signal frequency increases, the peak INL and DNL remain nearly unchanged until near the Nyquist input frequencies. The linearity is then dominated by the front-end pseudodifferential THAs. The dynamic performance of the converter is validated in the frequency domain as well. The frequency spectrum of the reconstructed signal is shown in Fig. 9, where the input signal frequency is about 941 MHz and the clock frequency is 2 GHz. The $0.5\ f_s - f_{\text{in}}$ tone is about 50 dB down, which implies the gain and timing errors between interleaved channels do not limit the linearity performance of the overall ADC system. The dominant harmonics (second and third) are contributed by the front-end pseudodifferential THAs. Fig. 10 depicts the measured spurious free dynamic range (SFDR) and signal-to-noise-and-distortion ratio (SNDR) versus the input signal frequency at 2-GHz sampling rate. At the low input frequency of 4 MHz, the SNDR and SFDR reach 36 and 48 dB, respectively. Near the Nyquist input frequencies (up to 941 MHz), the measured SNDR and SFDR remain above 30 and 35.5 dB. The analog input range is set to 1.0-V peak-to-peak differential. The input capacitance of the ADC is about 1 pF. Including the front-end THAs and on-chip

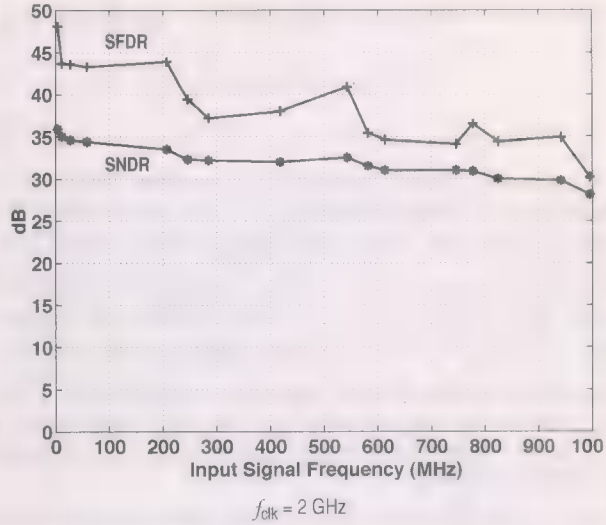


Fig. 10. Measured SNDR and SFDR as a function of input frequency.

clock buffers, the complete ADC consumes 310 mW of power from a single 1.8-V supply, while operating at 2-GHz conversion rate with input signal frequency up to 996 MHz.

V. Conclusion

A 2-GS/s 6-bit ADC with time-interleaving is demonstrated in a 0.18-μm 1P6M CMOS technology. A triple-cross connection method is invented to improve the offset averaging efficiency. Open-loop THAs with replica-based well-biasing are realized to ensure the dynamic performance up to Nyquist frequencies. This ADC is optimized to achieve a state-of-the-art figure-of-merit, defined as $(\text{Power})/(2^{\text{ENOB}} \cdot 2 \cdot \text{ERBW})$, of 3.5 pJ per conversion step.

Acknowledgment

The authors wish to thank Z. Wang for his assistance in chip layout and measurement.

References

[1] K. Kattmann and J. Barrow, "A technique for reducing differential nonlinearity errors in flash A/D converters," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 1991, pp. 170–171.

[2] K. Bult and A. Buchwald, "An embedded 240-mW 10-b 50-MS/s CMOS ADC in 1-mm^2," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1887–1895, Dec. 1997.

[3] M. Choi and A. A. Abidi, "A 6 b 1.3 Gsample/s A/D converter in 0.35 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 1847–1858, Dec. 2001.

[4] P. Scholtens and M. Vertregt, "A 6-b 1.6-Gsamples/s flash ADC in 0.18-μm CMOS using averaging termination," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1599–1609, Dec. 2002.

[5] J. White and A. Wilson Jr., "On the equivalence of spatial and temporal stability for translation invariant linear resistive networks," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 734–743, Sep. 1992.

[6] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 10, pp. 1433–1440, Oct. 1989.

[7] W. Black and D. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. SC-15, no. 12, pp. 1022–1029, Dec. 1980.

[8] B. Razavi, *Principles of Data Conversion System Design*. New York: IEEE Press, 1995.

[9] J. Doernberg, H. Lee, and D. Hodges, "Full-speed testing of A/D converter," *IEEE J. Solid-State Circuits*, vol. 19, no. 2, pp. 820–827, Feb. 1984.

A sinh Resistor and Its Application to tanh Linearization

Maziar Tavakoli, *Student Member, IEEE,* and Rahul Sarpeshkar, *Member, IEEE*

***Abstract*—We present a novel and simple subthreshold tunable resistor (sinh R) which exhibits a sinh I–V characteristic. This compact 8-transistor circuit generates an output current that is proportional to the sinh of its input differential voltage and has an offset-free characteristic, i.e., zero current at zero differential voltage, like a real resistor. In a 1.5-μm CMOS chip implementation, we achieved a common-mode rejection ratio (CMRR) of 46 dB. As an example application, we use the expansive properties of our sinh R to linearize the compressive properties of a tanh differential pair by degeneration and cancel all nonlinearities up to fifth order. We demonstrate good agreement between theory and experimental results.**

***Index Terms*—Distortion, filter, linearization techniques, sinh resistor, subthreshold operation, tanh differential pair.**

I. Introduction

RESISTORS with a sinh I–V characteristic could be useful in various nonlinear dynamical systems. For example, they can be used to implement attack times in automatic gain control circuits that quicken for larger input transients. To the best of our knowledge, a transistor-level implementation of a *tunable* sinh resistor has never been reported in the literature. We present a compact circuit that generates an output current proportional to the sinh of its input differential voltage.

Differential transconductors are essential elements in many analog electronic systems, such as filters, amplifiers, mixers, oscillators, and signal processing systems. Subthreshold differential pairs are attractive because of their low power consumption, large tuning range, and low transconductance, which allow them to efficiently implement low-frequency continuous-time filters; for example, in the audio range (20 Hz–20 kHz). Other applications for subthreshold circuits include biomedical implants, sensors and sensory networks, earthquake and vibration sensing, and low-power analog-to-digital (A/D) conversion. Since MOS transistors operating below threshold show an exponential I–V property, basic subthreshold differential pairs, like their bipolar counterparts, suffer from limited linear range and harmonic distortion produced by their tanh I–V transfer characteristic [1]–[3].

At the expense of a modest increase in area and power consumption, several linearizing schemes have been suggested in the literature to extend the input linear range of exponential (subthreshold MOS, bipolar) differential pairs, including source (emitter) degeneration via resistors [4], degeneration via diode-connected transistors [3], source degeneration via single or double diffusors (MOS transistors operating in the subthreshold ohmic region) [5], multiple parallel asymmetric differential pairs [6], [7], application of the input signal to the back-gate (well) terminals [1], gate degeneration [1], the use of a correlator or bump circuit [1], [8], or the combination of some of the above [9] or other [10]–[12] techniques. A larger linear range for a transconductor in thermal-noise limited cases translates to a rise in the dynamic range of filters built with such transconductors [1]. In this brief, we discuss how to use a sinh resistor to linearize a subthreshold differential pair by counteracting the compressive properties of a tanh with the expansive properties of a sinh to obtain a curve that is more linear than a tanh.

The outline of this brief is as follows. In Section II, we present the basic idea and the design of the sinh resistor and show data taken from a chip. We describe the implementation and the experimental results of a sinh-linearized tanh differential pair in Section III. We summarize in Section IV.

II. sinh Resistor (sinh R)

A. Basic Idea

Fig. 1(a) shows a two-port element with an expansive I–V characteristic. It is composed of a MOS transistor whose drain voltage is shifted up by V_a and coupled back to its gate terminal. To intuitively understand the operation of this element, we recognize that with a zero V_a, this element essentially acts like a diode. This means that when its voltage (V) is increased, its current (I) rises either in an exponential or a square-law manner (depending on its regime of operation), both of which are expansive. A tunable V_a allows for a tunable slope at the origin. The I–V curve is offset-free because zero drain-to-source voltage across a transistor always yields zero current. A sinh curve also has an (exponential) expansive quality, possesses a nonzero slope at the origin, and passes through the origin.

The latter two properties are also observed in the linear resistor of Fig. 1(b) and also in the compressive two-port element of Fig. 1(c), in which the gate-to-source voltage (V_{GS}) of a MOS transistor has been fixed at V_c. When the voltage across this element ($V_{DS} = V$) increases from zero, the current through this element (I) rises from zero, gradually flattens out (a compressive quality), and approaches its saturation value (I_{DSsat}) in an exponential or second-order fashion (based on its operation regime).

The element of Fig. 1(a) constitutes the basic core of our sinh resistor circuit. One problem with this element is that, unlike a typical resistor, it cannot function in a bi-directional way. One easy solution to implement a bi-directional sinh resistor is to place two expansive elements of Fig. 1(a) in parallel and opposite directions, as demonstrated in Fig. 2(a). However, a more elegant way to achieve bi-directionality, which shares bias voltage sources, is illustrated in Fig. 2(b). A single bias circuit determines which side is the drain and which side is the source and

Manuscript received March 15, 2004; revised August 10, 2004.

The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: maziar@mit.edu; rahuls@mit.edu).

Digital Object Identifier 10.1109/JSSC.2004.841015

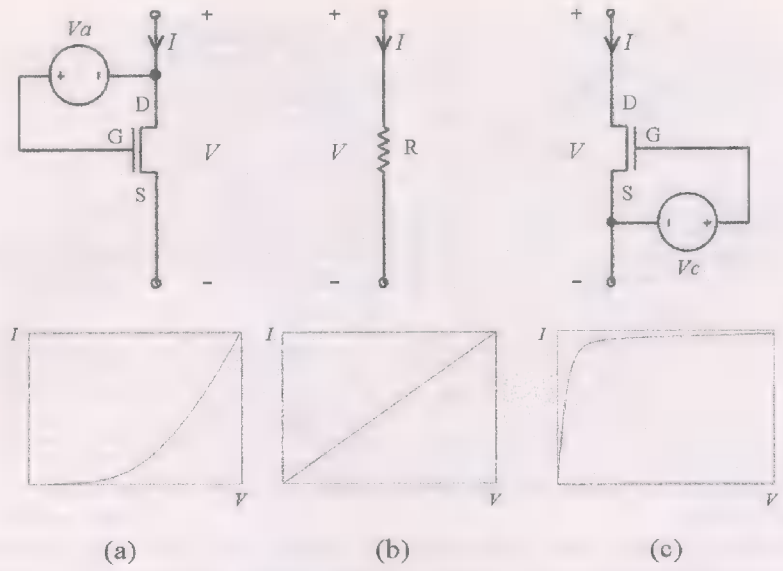


Fig. 1. Circuit representation and I–V characteristic of (a) an expansive, (b) a linear, and (c) a compressive two-port element.

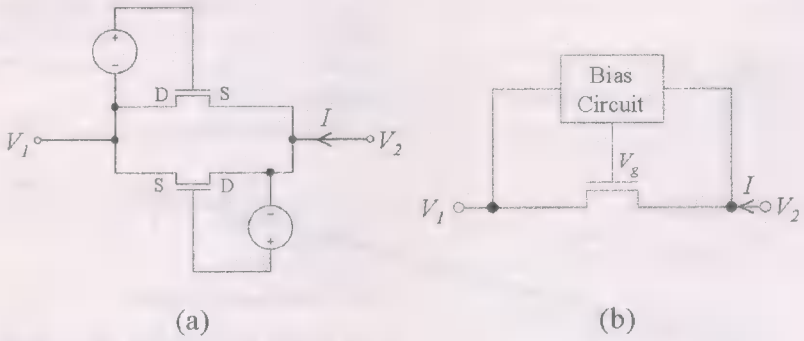


Fig. 2. Two different implementations of a bi-directional sinh resistor. (a) By the use of two expansive resistors in parallel and opposite directions. (b) By the use of a single bias circuit with source and drain inputs.

puts the appropriate voltage on the gate. If V_g is determined by the maximum of V_1 and V_2 (i.e., the drain), a sinh resistor is obtained. Similarly, if the minimum of V_1 and V_2 (i.e., the source) sets V_g, a tanh resistor is attained. Since tanh properties have been extensively realized by other methods in circuit design, the focus of the remainder of this brief will be on the sinh resistor.

B. Circuit Implementation

Fig. 3(a) shows the circuit schematic of a maximum circuit that can function as the bias circuit required in Fig. 2(b) to realize a sinh resistor. To analyze this circuit, we note that the current in a subthreshold MOS transistor is given by [2], [13]

$$I_{DS} = I_o e^{(\kappa V_{GB}/U_T)} \left(e^{(-V_{SB}/U_T)} - e^{(-V_{DB}/U_T)} \right) \quad (1)$$

$$\xrightarrow{\text{In Saturation: } V_{DS} \geq 5U_T}$$

$$I_{DS} = I_o e^{((\kappa V_{GB} - V_{SB})/U_T)} \quad (2)$$

where V_{GB}, V_{SB}, and V_{DB} are the gate-to-body, source-to-body, and drain-to-body voltages, respectively; κ is the subthreshold exponential coefficient; I_o is the subthreshold current-scaling parameter; and $U_T = kT/q$ is the thermal voltage (about 25.9 mV at room temperature). In the simple model of (1), the effect of a nonzero drain-to-source conductance on I_{DS} has been ignored.

In the circuit of Fig. 3(a), if we ignore the output resistance of the top T_3-T_4 pMOS current mirror, we can write (note that the bodies of all nMOS transistors in our n-well process are tied to the substrate, which is connected to ground; i.e., $V_B = 0$ V)

$$I_{DS1} + I_{DS2} = I_{DS5} = \frac{I_b}{2}$$

$$\Rightarrow e^{(\kappa V_1/U_T)} + e^{(\kappa V_2/U_T)} = e^{(\kappa V_{\text{out}}/U_T)}$$

$$\Rightarrow V_{\text{out}} = \frac{U_T}{\kappa} \ln(e^{(\kappa V_1/U_T)} + e^{(\kappa V_2/U_T)}). \quad (3)$$

If one of the input signals is much larger than the other one, (3) simplifies to $V_{\text{out}} = \max(V_1, V_2)$. In a similar approach, a pMOS version of this maximum circuit forms a minimum circuit that can be used to create a tanh resistor with the topology of Fig. 2(b).

Fig. 3(b) illustrates the circuit schematic of our sinh resistor (sinh R) based on the implementation idea of Fig. 2(b) which employs the maximum circuit of Fig. 3(a). The voltage V is equal to the maximum of V_1 and V_2, that is, the drain side of the main sinh transistor T_8 in Fig. 3(b). V is shifted up by a diode drop to set V_g which is then connected to the gate terminal of T_8.

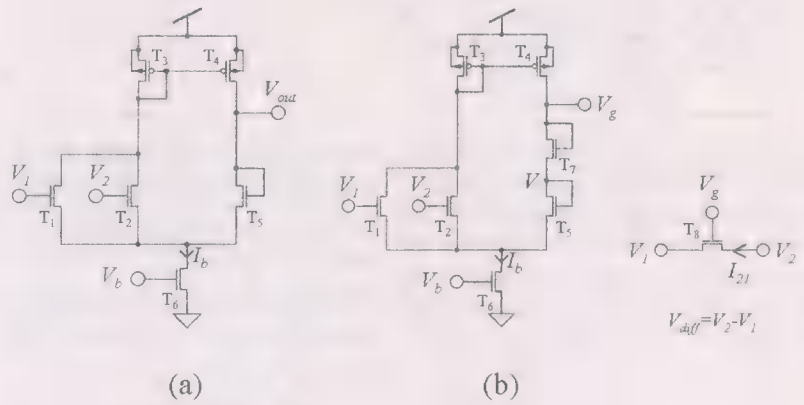


Fig. 3. Circuit schematic of (a) the maximum circuit and (b) the sinh resistor (sinh R).

Thus, the expanding element of Fig. 1(a) has been successfully replicated, and as we show below, the body effect of T_8 is also compensated for.

For the sinh R circuit, we can write

$$I_{DS7} = \frac{I_b}{2} \Rightarrow I_o e^{((\kappa V_g - V)/U_T)} = \frac{I_o}{2} e^{(\kappa V_b/U_T)} \Rightarrow \frac{\kappa V_g - V}{U_T} = \frac{\kappa V_b}{U_T} - \ln 2 \quad (4)$$

$$I_{\sinh} = I_{21} = I_o e^{(\kappa V_g/U_T)} (e^{(-V_1/U_T)} - e^{(-V_2/U_T)}). \quad (5)$$

Equation (5) can be further simplified to

$$I_{\sinh} = I_{21} \overset{(4)}{=} \frac{I_b}{2} e^{(V/U_T)} \times (e^{(-V_1/U_T)} - e^{(-V_2/U_T)}) \xrightarrow{(3),\ V=V_{\text{out}}}$$

$$I_{\sinh} = I_{21} = \frac{I_b}{2} \left[e^{(\kappa V_1/U_T)} + e^{(\kappa V_2/U_T)} \right]^{1/\kappa} \times (e^{(-V_1/U_T)} - e^{(-V_2/U_T)}). \quad (6)$$

For analytical purposes, we decompose the two inputs to our sinh R as $V_1 = V_{CM} - V_{\text{diff}}/2$ and $V_2 = V_{CM} + V_{\text{diff}}/2$. Performing some algebra on (6) yields

$$I_{\sinh} = I_{21} = I_o' \sinh\left(\frac{V_{\text{diff}}}{2U_T}\right) \text{ where } I_o' = I_b \left[2\cosh\left(\frac{\kappa V_{\text{diff}}}{2U_T}\right) \right]^{1/\kappa}. \quad (7)$$

Assuming κ is close to one, we can approximate (7) to obtain

$$I_{\sinh} = I_{21} \approx I_b \left[2\cosh\left(\frac{V_{\text{diff}}}{2U_T}\right) \sinh\left(\frac{V_{\text{diff}}}{2U_T}\right) \right] = I_b \sinh\frac{V_{\text{diff}}}{U_T}. \quad (8)$$

Viewing the entire circuit of Fig. 3(b) as a two-port element, (8) shows that the current through this element (I_{21}) is proportional to the sinh of the differential voltage applied to it ($V_{\text{diff}} = V_2 - V_1$); thus, we have created a resistor with a sinh I–V characteristic. Note that there is no current when V_{diff} is zero. The absence of V_{CM} in (8) implies that as long as V_{diff} is fixed, the current has no dependence on common-mode voltage or on the body effect of T_8, just like in a real resistor. An intuitive way to understand the common-mode rejection is that, for a fixed differential voltage of ΔV between the drain and the source of T_8, its current only depends on its $\kappa V_g - V_s$ or $\kappa V_g - V_d$. However, since V_g and V_d are connected to the maximum circuit, no matter what their common-mode voltage level is, $\kappa V_g - V_d = \kappa V_g - V$ is set by $I_b/2$, the current through transistor T_7 [see (4)].

The transconductance (g_m) of our sinh R is given by

$$g_m = \frac{dI_{21}}{dV_{\text{diff}}} = \frac{I_b}{U_T} \cosh\left(\frac{V_{\text{diff}}}{U_T}\right)\Big|_{V\text{diff}=0} = \frac{I_b}{U_T} \frac{A}{V}. \quad (9)$$

C. Experimental Results

A circuit prototype of our sinh R, as illustrated in Fig. 3(b), was fabricated in a 1.5-μm CMOS MOSIS n-well process. All the transistors had the same size (4.8 μm/4.8 μm). The experimental tests were all run on a 5-V power supply. The common-mode voltage of the signals applied to the sinh R was 2.5 V, unless otherwise stated.

The current versus differential voltage (V_{diff}) characteristic of our sinh R is plotted in Fig. 4(a) for three different values of bias voltage V_b equal to 0.35 V, 0.40 V, and 0.45 V corresponding to 110 pA, 470 pA, and 1.75 nA of bias current I_b, respectively. A magnified view of the curves in Fig. 4(a) near the origin is shown in Fig. 4(b). The theoretical fits have been calculated using (8) for Fig. 4(a) and (9) for Fig. 4(b). We see that the experimental data are in good agreement with theory. However, since κ is less than one in practice, the sinh current formula of (8) and also the transconductance formula of (9) slightly underestimate the actual $I_{\sinh}$ and g_m. At large magnitudes of V_{diff} (not shown in these figures), the theoretical current eventually surpasses the experimental results because T_8 gradually leaves the subthreshold exponential region and operates in the above-threshold square-law regime.

Fig. 5 demonstrates the variation of sinh current with the common-mode dc voltage (V_{CM}), at a fixed V_{diff}. We observe that for a V_{CM} range of 3 V (0.5–3.5 V), the current changes only by a factor of 1.8. To compare, we note that such a variation in current could have been caused by a change of only about 15 mV in V_{diff}. In other words, the effect of common-mode

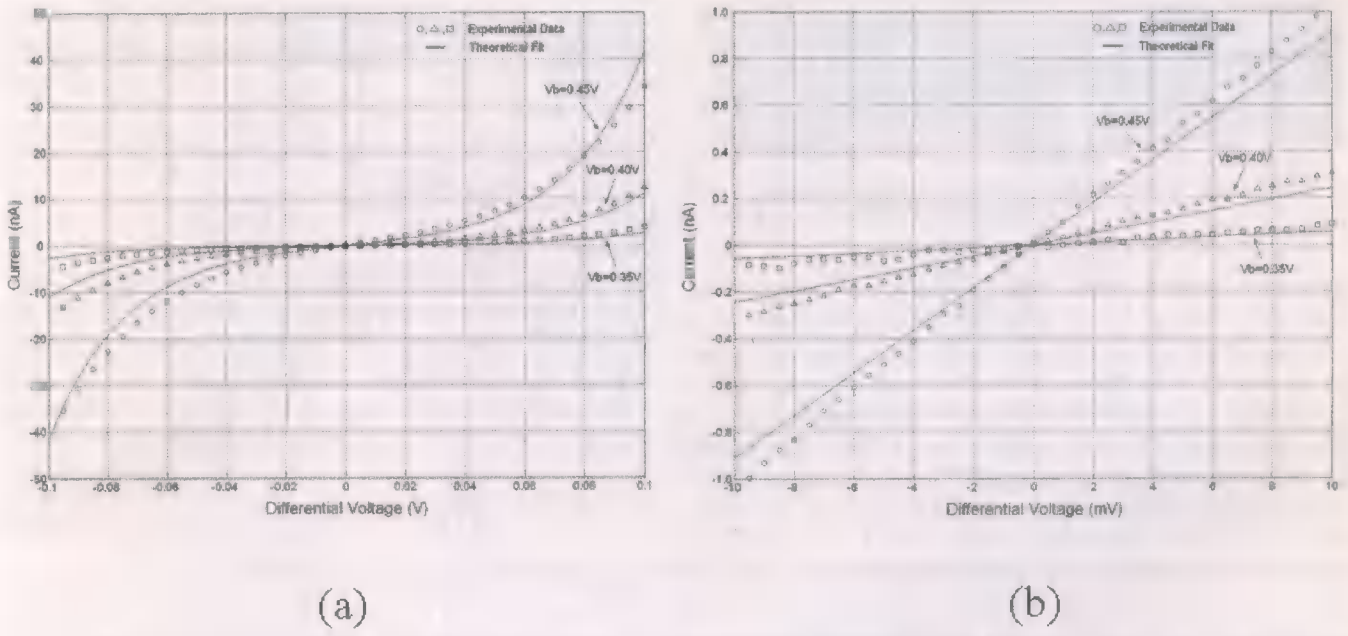


Fig. 4. (a) Current versus differential voltage characteristic of the sinh R shown in Fig. 3(b) and (b) its magnified view near the origin.

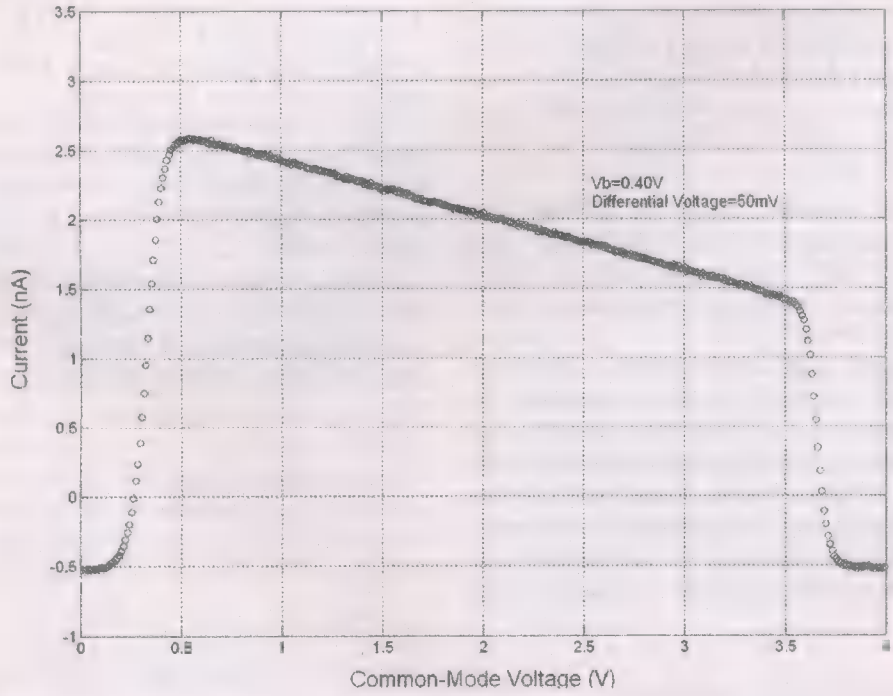


Fig. 5. Common-mode characteristic of the sinh R shown in Fig. 3(b) at a fixed differential voltage.

voltage on the current is about 200 times weaker than the effect of differential voltage, translating to a common-mode rejection ratio (CMRR) of 46 dB. The small variation of current with V_{CM} seen in Fig. 5, which is not predicted by (7) or (8), is due to the fact that κ slightly rises with an increase in the common-mode voltage [1] which causes the sinh current to drop. The nonzero drain-to-source conductances of transistors also have an effect. The sudden drop of the current at two ends of Fig. 5 is due to transistors T_4 and T_6 in the maximum circuit of Fig. 3(a) coming out of saturation, thus disrupting the proper function of the circuit.

III. Example Application in a Differential Pair

A. Basic Idea and Circuit Implementation

The circuit schematic of a standard CMOS source-coupled differential pair is shown in Fig. 6(a). Although this transconductor uses additional current mirrors to achieve a wide output voltage range [2], the presence of this extra circuitry does not affect the arguments presented in this section regarding differential pairs, assuming ideal mirrors. If the transistors are biased in subthreshold regime, the differential output current (I_{out})

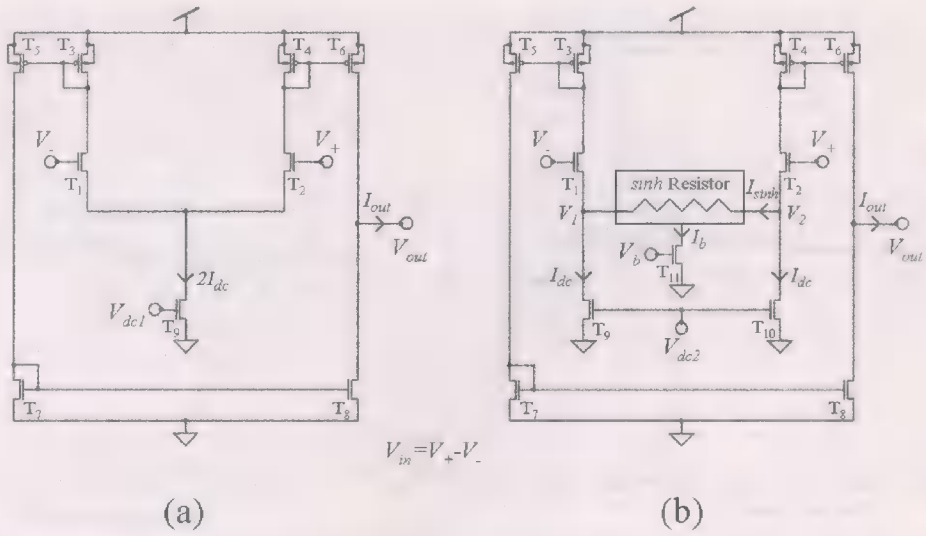


Fig. 6. Circuit schematic of (a) basic CMOS differential pair with wide output voltage range and (b) a sinh-linearized CMOS differential pair.

and the transconductance (G_m) of this circuit are shown to be [1]–[3]

$$I_{\text{out}} = 2I_{dc}\tanh\left(\frac{V_{\text{in}}}{V_L}\right) = 2I_{dc}\tanh\left(\frac{\kappa V_{\text{in}}}{2U_T}\right)$$
$$\Rightarrow G_m = \left.\frac{dI_{\text{out}}}{dV_{\text{in}}}\right|_{V_{\text{in}}=0} = \frac{2I_{dc}}{V_L} = \frac{\kappa(2I_{dc})}{2U_T}\frac{A}{V}. \quad (10)$$

tanh is a nonlinear function, which can produce harmonic distortion in the signal. Also, the linear range ($V_L = 2U_T/\kappa \approx \pm 75$ mV) of this transconductor is too small for many applications where it is desirable to handle large inputs without distortion.

One intuitive solution to improve the linearity problem of such a tanh differential transconductor is to compensate for the compressive properties of a tanh with the expansive properties of a nonlinear function of its own hyperbolic kind, such as a sinh, to obtain a more linearized curve compared to a pure tanh. To this end, we simply source degenerate our differential pair with the sinh R developed in Section II. The circuit of such a sinh-degenerated CMOS differential pair is demonstrated in Fig. 6(b).

B. Theoretical Analysis and Experimental Results

Using (2) for transistors T_1 and T_2 in the circuit of Fig. 6(b) and calculating their current ratio, sum, and difference, we derive

$$\frac{I_2}{I_1} = e^{((\kappa V_{\text{in}}/U_T)-((V_2-V_1)/U_T))} \Rightarrow \frac{I_2 - I_1}{I_2 + I_1}$$
$$= \frac{2I_{\sinh}}{2I_{dc}} = \frac{I_{\text{out}}}{2I_{dc}}$$
$$= \tanh\left(\frac{\kappa V_{\text{in}}}{2U_T} - \frac{V_2 - V_1}{2U_T}\right)$$
$$\Rightarrow \tanh^{-1}\left(\frac{I_{\text{out}}}{2I_{dc}}\right) + \frac{V_2 - V_1}{2U_T} = \frac{\kappa V_{\text{in}}}{2U_T}. \quad (11)$$

Applying (8) to the sinh R element of Fig. 6(b), (11) is transformed to

$$\tanh^{-1}\left(\frac{I_{\text{out}}}{2I_{dc}}\right) + \frac{1}{2}\sinh^{-1}\left(\frac{I_{\text{out}}}{2I_b}\right) = \frac{\kappa V_{\text{in}}}{2U_T} \quad (12)$$

In the compact formula of (12), it is reassuring to recognize that: first, with no $\sinh^{-1}$ term (or equivalently $I_{dc} \ll I_b$), (12) reduces to (10) which is the characteristic of a basic differential pair, as expected; second, since the argument of the $\tanh^{-1}$ function implies $|I_{\text{out}}| \leq 2I_{dc}$, $2I_{dc}$ sets the limiting current of the sinh-linearized differential transconductor in a similar fashion as it does for the basic differential pair of Fig. 6(a).

To study the linearity of the transfer curve of (12), let us consider the Taylor expansions of the following functions about zero

$$\tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$$
$$= x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \frac{1}{7}x^7 + \cdots \quad |x| \leq 1,$$
$$\sinh^{-1}(y) = \ln\left(y + \sqrt{1+y^2}\right)$$
$$= y - \frac{1}{6}y^3 + \frac{3}{40}y^5 - \frac{5}{112}y^7 + \cdots \quad (13)$$

To obtain a maximally linear I_{out}–V_{in} curve, we can Taylor expand (12) based on (13) and then adjust I_{dc} and I_b to eliminate the cubic-distortion term. This is achieved if

$$\frac{1}{24I_{dc}^3} - \frac{1}{96I_b^3} = 0 \Rightarrow I_b = \frac{I_{dc}}{\sqrt[3]{4}} \approx \frac{I_{dc}}{1.59} \quad (14)$$

If the optimal condition of (14) is satisfied, the first remaining nonlinearity will be due to the fifth-order term and the Taylor expansion of (12) reduces to

$$1.794x + 0.578x^5 - 0.424x^7 + \cdots = \frac{\kappa V_{\text{in}}}{2U_T} \quad (15)$$

with $x = I_{\text{out}}/2I_{dc}$. In comparison, the I_{out}–V_{in} characteristic of a simple differential transconductor [see (10)] has cubic distortion

$$\tanh^{-1}\left(\frac{I_{\text{out}}}{2I_{dc}}\right) = \frac{\kappa V_{\text{in}}}{2U_T} \Rightarrow x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \frac{1}{7}x^7 + \cdots = \frac{\kappa V_{\text{in}}}{2U_T}. \quad (16)$$

For example, at $x = 0.5$, the tanh differential pair has cubic distortion of 8.3%, as compared to the sinh-linearized differential transconductor which has only fifth-order distortion of about 2%. Therefore, the tanh is made more linear by sinh degeneration.

From (12), the transconductance (G_m) of our new differential pair is found to be

$$G_m = \left(\left.\frac{dV_{\text{in}}}{dI_{\text{out}}}\right|_{V_{\text{in}}=I_{\text{out}}=0}\right)^{-1} = \frac{2I_{dc}}{V_L} = \frac{\kappa}{2U_T}\frac{1}{\frac{1}{2I_{dc}}+\frac{1}{4I_b}} = \frac{\kappa\,(2I_{dc})}{2U_T}\frac{1}{1+\frac{I_{dc}}{2I_b}}\frac{A}{V}. \quad (17)$$

Compared to (10), (17) suggests that the G_m is decreased and thus the V_L is increased (remember that the maximum current remains the same at $2I_{dc}$) by a factor equal to $1 + I_{dc}/2I_b$. For the optimal case of (14), this factor is almost 1.8 (seen also in (15)), which makes the new V_L equal to ± 135 mV. This 80% increase in linear range costs a 16% (i.e., $I_b/4I_{dc}$) increase in power consumption.

CMOS differential pairs without and with sinh-linearization were fabricated in a 1.5-μm CMOS chip with the same size for all the transistors (4.8 μm/4.8 μm). Fig. 7 shows a photograph of the chip. The experimental output current versus input voltage dc characteristics for these two circuits are plotted in Fig. 8(a). The data were taken with $2I_{dc} = 10$ nA. The optimal condition of (14) was also satisfied in the second circuit. We clearly see that the curve with sinh linearization has a smaller slope (transconductance) and a larger linear range than the one without. The observed improvement factor is 1.7, close to 1.8 that theory predicts.

To further study the linearity of the transfer curves, which is difficult to examine visually from the graphs in Fig. 8(a), we performed the following analysis: Having fixed $2I_{dc}$ at 10 nA, we varied I_b over a large range. For each setting, we measured the voltage (normalized by $2U_T/\kappa$) versus current (normalized by $2I_{dc}$) transfer characteristic of the sinh-linearized differential pair. We fit a fifth-order polynomial to each experimental curve. In other words, we experimentally derived the polynomial approximation to the main formula of (12) for different I_b's. In Fig. 8(b), we plot the magnitudes for the coefficient of the 1st (linear) term and the coefficient of the 3rd (cubic) term as I_{dc}/I_b is changed. We see that the minimum magnitude for the cubic term occurs at $I_{dc}/I_b = 1.72$, close to theoretical value of 1.59 predicted in (14).

We also configured the transconductors of Fig. 6 as two simple first-order low-pass G_m-C filters (i.e., output terminal connected to the negative input terminal and a capacitor) with cutoff frequencies (i.e., $G_m/2\,\pi C$) of 4 kHz. The measured frequency response of the filter with sinh degeneration is illustrated in Fig. 9. As another experimental test of our sinh-linearizing idea, we applied a $280mV_{\text{pp}}$ ($100mV_{\text{rms}}$) passband sinusoidal signal input at 110 Hz to these filters. We measured the spectrum of their output signals with an SR785 Spectrum Analyzer. We observed that the rms amplitude of the third harmonic in the sinh-linearized filter output is smaller by a factor of 27 (28.6 dB) than the same term in the standard tanh filter output. In fact, in our filters, the second harmonic was the main contributor to nonlinearity and the total harmonic distortion was measured at about 1%. The presence of the second harmonic is attributed to device mismatches (among our relatively small transistors), variations in κ with voltage, and the existence of parasitic capacitors on the common source nodes [14], which all distort the symmetry of the I–V transfer curve. As is well known, employing a fully differential G_m-C filter topology significantly reduces even-order nonlinearities [15]. In such a circuit, the effect of our sinh-linearizing scheme on improving nonlinearity would be substantial.

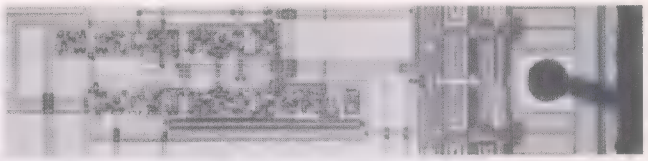


Fig. 7. Microphotograph of the chip containing the circuits of the sinhR and the basic and sinh-linearized differential pairs.

We also briefly discuss the noise of our sinh R and sinh-linearized transconductor. Noise is important because it determines the lower bound on the dynamic range. For low subthreshold current levels, the $1/f$ noise of transistors is usually negligible compared to thermal noise [1]. In the circuit of Fig. 3(b), the current noise of the sinh R is generated only by the shot noise of the main sinh transistor T_8 if V_{DS} is zero. For nonzero V_{DS}, the noise of the maximum circuit multiplicatively modulates the current flowing through T_8 and is thus V_{DS} dependent. When both inputs are (small-signal) grounded, the current noise power spectral density of T_8 is $4qI_{DSsat}$ [13], where I_{DSsat} is the saturation current of the transistor, given by (2). Therefore

$$\overline{i^2_{n\,\sinh}} = 4qI_{DSsat} \overset{(2)}{=} 4qI_o e^{((\kappa V_g - V_1)/U_T)} \xrightarrow{(3),\ V_1=V_2} \overline{i^2_{n\,\sinh}} = 4qI_o e^{((\kappa V_g - (V - \ln 2\times U_T/\kappa))/U_T)} \overset{(4)}{=} 4q\frac{I_b}{2}2^{(1/\kappa)} = 2^{1+(1/\kappa)}qI_b \overset{\kappa\simeq 1}{=} 4qI_b\frac{A^2}{H_z}. \quad (18)$$

Thus, in this case, the input-referred voltage noise of the sinh-linearized transconductor of Fig. 6(b) is found by a standard procedure [1], [16] to be

$$\overline{v^2_{\text{in}}} = \frac{8(2qI_{dc}) + 4\overline{i^2_{n\,\sinh}}}{G_m^2} = \frac{16qI_{dc}\left(1+\frac{I_b}{I_{dc}}\right)}{G_m^2}\frac{V^2}{H_z} \quad (19)$$

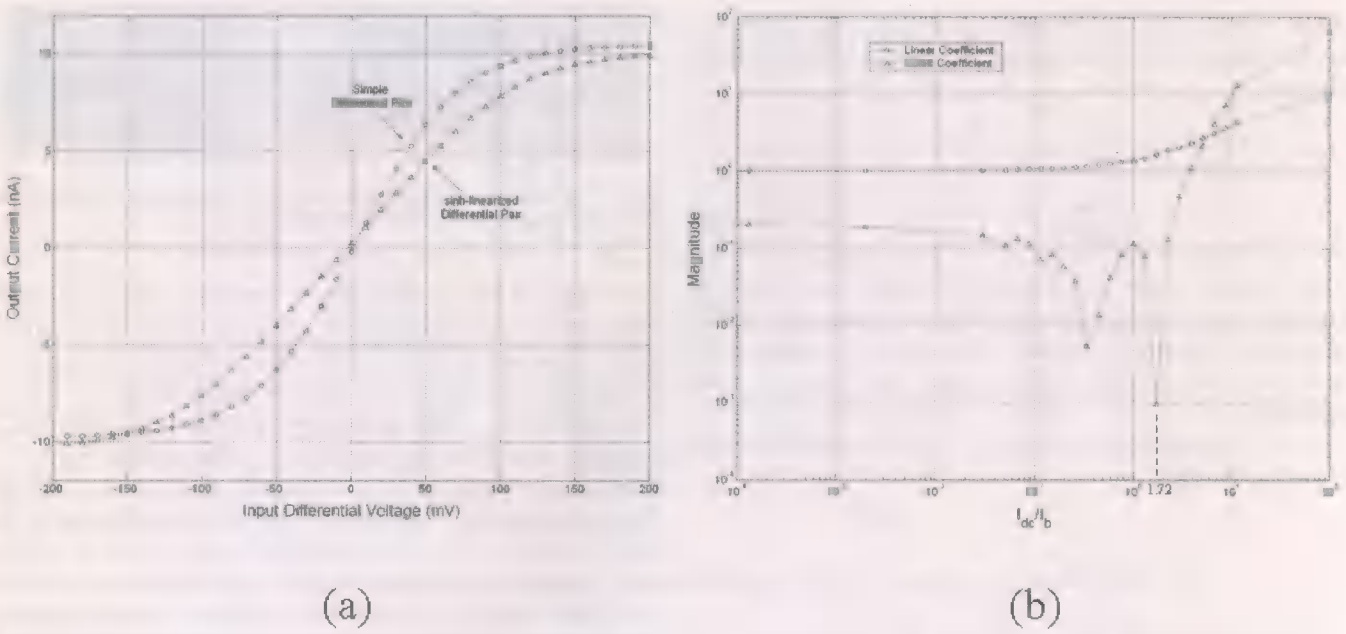


Fig. 8. (a) Experimental output current versus input voltage characteristics of the two circuits displayed in Fig. 6. (b) Magnitudes of polynomial coefficients that are fit to measured I–V curves of the transconductor of Fig. 6(b) for different values of I_{dc}/I_b.

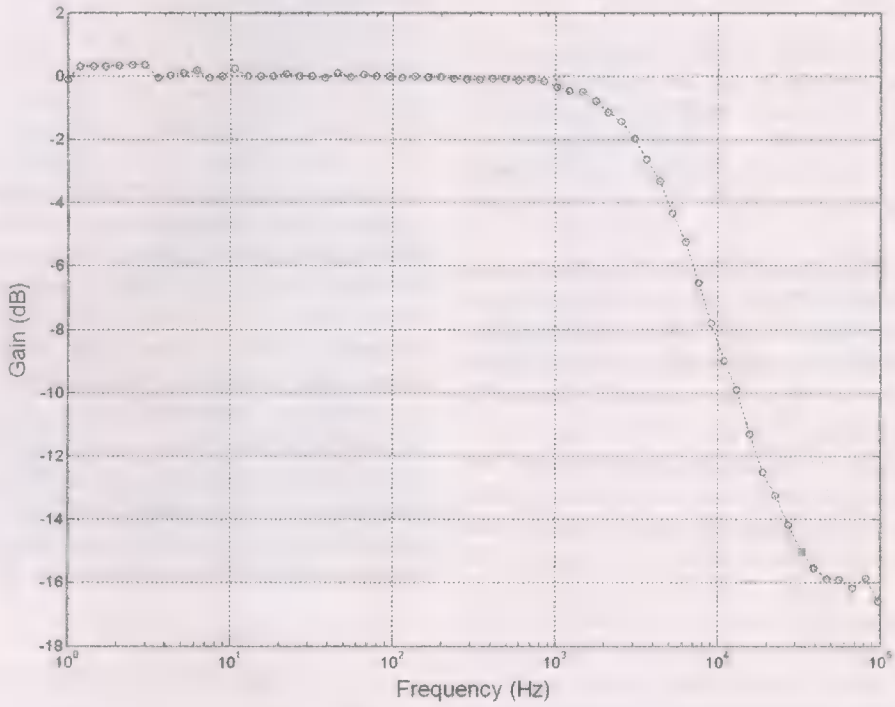


Fig. 9. Experimental frequency response of the sinh-linearized transconductor of Fig. 6(b) configured as a simple first-order low-pass G_m-C filter.

where G_m is given by (17). With $2I_{dc} = 10$ nA and I_b/I_{dc} ratio set according to the optimal condition of (14), the input-referred voltage noise was theoretically calculated to be 1.9 $\mu\text{V}/\sqrt{\text{Hz}}$, and was measured at about 1.7 $\mu\text{V}/\sqrt{\text{Hz}}$.

Resistive [4] and diode [3] degeneration are among the two most widely used linearization techniques in circuit design. The main shortcoming of resistive degeneration besides the impracticality of creating large passive resistors required in circuits operating in subthreshold is that a resistor, as a linear element, has limited ability to oppose and improve the distortion introduced by inherently nonlinear exponential elements like transistors. A diode-degenerated differential pair also suffers from the same deficiency and essentially produces the same level of distortion as a simple differential pair does. The $\sinh R$, on the other hand, can exploit its own nonlinearity in a wise way to counteract and cancel unwanted nonlinearities. In Table I, we compare some of the characteristics of a basic, a resistive-degenerated, and a diode-degenerated $\tanh$ differential pair with those of our $\sinh$-degenerated transconductor. We see that an important advantage of our scheme is that it can be utilized to

TABLE I
CHARACTERISTICS OF BASIC AND VARIOUS DEGENERATED tanh DIFFERENTIAL PAIRS

Transconductor Characteristics	Comments	Basic Differential Pair	Resistive-degenerated Differential Pair	Diode-degenerated Differential Pair	*Sinh*-degenerated Differential Pair
Linear Range (V_L)	Normalized to $2U_T/\kappa$ (V)	1	1.8	2.4	1.8
Cubic (3rd) Harmonic Distortion	@ $I_{out}/2I_{dc}$=0.5	8.3%	4.6%	8.3%	0%
Total Harmonic Distortion (THD)	@ $I_{out}/2I_{dc}$=0.5	8.4%	4.65%	8.4%	2%
Power	Normalized to $4I_{dc}V_{dd}$ (W)	1	1	1	1.16
Area/Number of Transistors	——	9	Very Large!	11	18
Effective Number of Noise-contributing Transistors (N)	——	8	13	10	13
Input dc Voltage Range	——	>0.65V	>0.67V	>1.15V	>1.15V

Notes: $2I_{dc}$=10nA
I_b=3.1nA (to satisfy the optimal condition of (14))
R=8.2MΩ (to set the same linear range for both resistive-degenerated and *sinh*-degenerated differential pairs)

eliminate cubic distortion, a useful feature that can never be achieved by resistive or diode degeneration or even *most* of the other linearization schemes introduced in Section I. This quality results in a lower total harmonic distortion (THD) and a more linear I–V curve for the transconductor. However, we should note that, like every other engineering approach, our technique that has been optimized for minimal harmonic distortion does not necessarily exhibit the best performance in all the other relevant properties, as we observe in Table I. Our scheme can, thus, be used in the design of transconductor circuits in which minimal distortion is of paramount interest.

IV. CONCLUSION

We described the basic idea and a compact CMOS implementation of a tunable resistor that possesses a sinh I–V characteristic. We showed that the current of such a resistor depends only on the sinh of its input differential voltage, not on its common-mode value, just like a normal resistor. We presented and justified experimental results that were in good agreement with our theoretical predictions. As an example application, we utilized our sinh R to degenerate a compressive subthreshold tanh differential pair and adjusted the circuit to cancel the cubic distortion introduced by a pure tanh curve which effectively widens the linear range by 80%. We also confirmed the effectiveness of our linearization technique in a first-order G_m-C filter where we reduced the third harmonic distortion by a factor of 27. The achieved extra linearity and its consequent drop in distortion are desirable qualities in many applications for differential transconductors, such as filters, mixers, and amplifiers.

REFERENCES

[1] R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A low-power wide-linear-range transconductance amplifier," *Analog Integrat. Circuits Signal Process.*, vol. 13, no. 1/2, pp. 123–151, May/Jun. 1997.

[2] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.

[3] L. Watts, D. A. Kerns, R. F. Lyon, and C. A. Mead, "Improved implementation of the silicon cochlea," *IEEE J. Solid-State Circuits*, vol. 27, no. 5, pp. 692–700, May 1992.

[4] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. New York: Wiley, 2001.

[5] P. M. Furth and A. G. Andreou, "Linearised differential transconductors in subthreshold CMOS," *Electron. Lett.*, vol. 31, no. 7, pp. 545–547, 1995.

[6] H. Tanimoto, M. Koyama, and Y. Yoshida, "Realization of a 1-V active filter using a linearization technique employing plurality of emitter-coupled pairs," *IEEE J. Solid-State Circuits*, vol. 26, no. 7, pp. 937–945, Jul. 1991.

[7] P. M. Furth and H. A. Ommani, "Low-voltage highly-linear transconductor design in subthreshold CMOS," in *Proc. 40th Midwest Symp. Circuits and Systems*, vol. 1, Aug. 1997, pp. 156–159.

[8] T. Delbruck, "Bump circuits for computing similarity and dissimilarity of analog voltages," *Caltech Computation Neural Syst. Memo*, no. 26, May 1993.

[9] R. R. Harrison, "A wide-linear-range subthreshold CMOS transconductor employing the back-gate effect," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 3, May 2002, pp. 727–730.

[10] G. Wilson, "Linearised bipolar transconductor," *Electron. Lett.*, vol. 28, no. 4, pp. 390–391, 1992.

[11] M. T. Abuelma'atti, "Universal CMOS current-mode analog function synthesizer," *IEEE Trans. Circuits Syst. I*, vol. 49, no. 10, pp. 1468–1474, Oct. 2002.

[12] J. W. Fattaruso and R. G. Meyer, "MOS analog function synthesis," *IEEE J. Solid-State Circuits*, vol. 22, no. 12, pp. 1056–1063, Dec. 1987.

[13] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. Boston, MA: McGraw-Hill, 1999.

[14] E. K. De Lange, O. De Feo, and A. Van Staveren, "Modeling differential pairs for low-distortion amplifier design," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 1, May 2003, pp. 261–264.

[15] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*. New York: Wiley, 1997.

[16] M. Tavakoli and R. Sarpeshkar, "An offset-canceling low-noise lock-in architecture for capacitive sensing," *IEEE J. Solid-State Circuits*, vol. 38, no. 2, pp. 244–253, Feb. 2003.

An Ultra-Wideband CMOS Low Noise Amplifier for 3–5-GHz UWB System

Chang-Wan Kim, Min-Suk Kang, Phan Tuan Anh, Hoon-Tae Kim, and Sang-Gug Lee

***Abstract*—An ultra-wideband (UWB) CMOS low noise amplifier (LNA) topology that combines a narrowband LNA with a resistive shunt-feedback is proposed. The resistive shunt-feedback provides wideband input matching with small noise figure (NF) degradation by reducing the Q-factor of the narrowband LNA input and flattens the passband gain. The proposed UWB amplifier is implemented in 0.18-μm CMOS technology for a 3.1–5-GHz UWB system. Measurements show a −3-dB gain bandwidth of 2–4.6 GHz, a minimum NF of 2.3 dB, a power gain of 9.8 dB, better than −9 dB of input matching, and an input IP3 of −7 dBm, while consuming only 12.6 mW of power.**

***Index Terms*—Broadband, CMOS, feedback, low noise amplifier, RF, ultra-wideband.**

I. Introduction

RECENTLY, the interest in ultra-wideband (UWB) system for wireless personal area network (WPAN) application has increased significantly, though the international standard has yet to be finalized. The allocated frequency band of the UWB system is 3.1–10.6 GHz (low-frequency band: 3.1–5 GHz; high-frequency band: 6–10.6 GHz). Two recent major proposals [1], [2] for the IEEE 802.15.3a propose that data rates of up to 400–480 Mb/s can be obtained using only the low-frequency band. The low-frequency band has been allocated for the development of the first-generation UWB system. CMOS technology is a satisfactory choice for the implementation of the low band UWB system when considering the time to market, hardware cost, the degree of difficulty, etc.

Until now, reported CMOS-based wideband amplifiers tend to be dominated by two different topologies: the distributed and resistive shunt-feedback amplifiers. The distributed amplifiers [3], [4] normally provide wide bandwidth characteristics but tend to consume large dc current due to the distribution of multiple amplifying stages, which makes them unsuitable for low-power application. The resistive shunt-feedback-based amplifiers [5], [6] provide good wideband matching and flat gain, but tend to suffer from poor noise figure (NF) and large power dissipation. In the resistive shunt-feedback amplifier, input resistance is determined by the feedback resistance divided by the loop-gain of the feedback amplifier [7]. Therefore, the feedback resistor tends to be a few hundred ohms in order to match the low signal source resistance of typically 50 Ω, leading to significant NF degradation. Furthermore, even with a moderate amount of voltage gain, the amplifier requires a rather large amount of current, especially in the CMOS, due to its strong dependence for voltage gain on the transconductance of the amplifying transistor. Recently, a new topology of a wideband amplifier for UWB system, which adopts a bandpass LC filter at the input of the cascode low noise amplifier (LNA) for wideband input matching, has been reported in [8] and [9]. The bandpass filter-based topology incorporates the input impedance of the cascode amplifier as a part of the filter, and shows good performances while dissipating small amounts of dc power. However, the adoption of the LC filter at the input mandates a number of reactive elements, which could lead to a larger chip area and NF degradation in the case of on-chip implementation, or the additional external components.

This paper proposes a new low power, low noise, and wideband amplifier combining a narrowband LNA with the conventional resistive shunt-feedback. The design principles and the measurement results of the implemented 3.1–5-GHz UWB LNA are described.

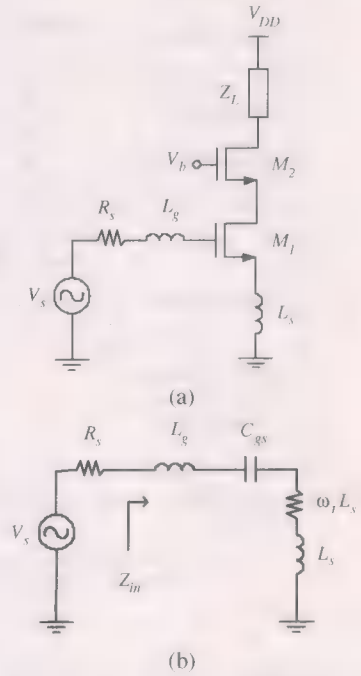


Fig. 1. Narrowband LNA topology. (a) Overall schematic. (b) Small-signal equivalent circuit at the input.

II. Design of Wideband Amplifier

Fig. 1(a) shows a typical narrowband cascode LNA topology. In Fig. 1(a), the inductor L_s is added for simultaneous noise and input matching and L_g for the impedance matching between the source resistance R_s and the input of the LNA [10]. Fig. 1(b) shows the small-signal equivalent circuit for the input part of the overall LNA, where C_{gs} represents the gate-source capacitance of the input transistor M_1. In Fig. 1(b), a series combination

Manuscript received April 8, 2004; revised August 26, 2004.

C.-W. Kim, M.-S. Kang, P. T. Anh, and S.-G. Lee are with the Information and Communications University, Yuseong, Daejeon, 305-600, Korea (e-mail: cwkim@icu.ac.kr).

H.-T. Kim is with the Samsung Advanced Institute of Technology, Suwon 440-600, Korea.

Digital Object Identifier 10.1109/JSSC.2004.840951

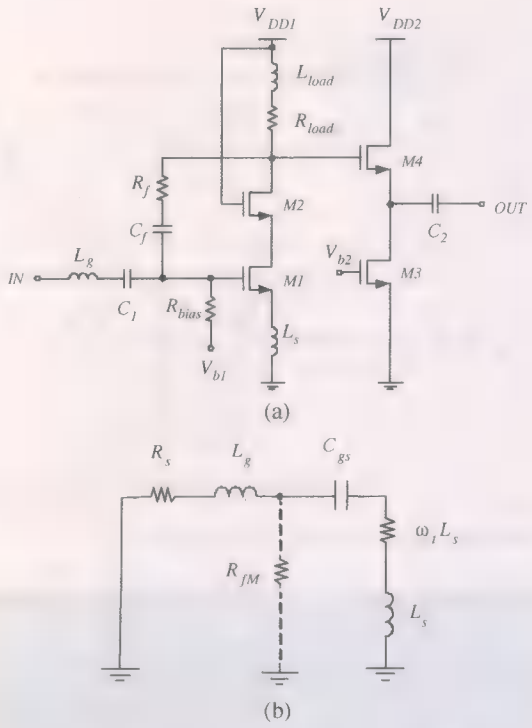


Fig. 2. UWB LNA topology. (a) Overall schematic. (b) Small-signal equivalent circuit at the input.

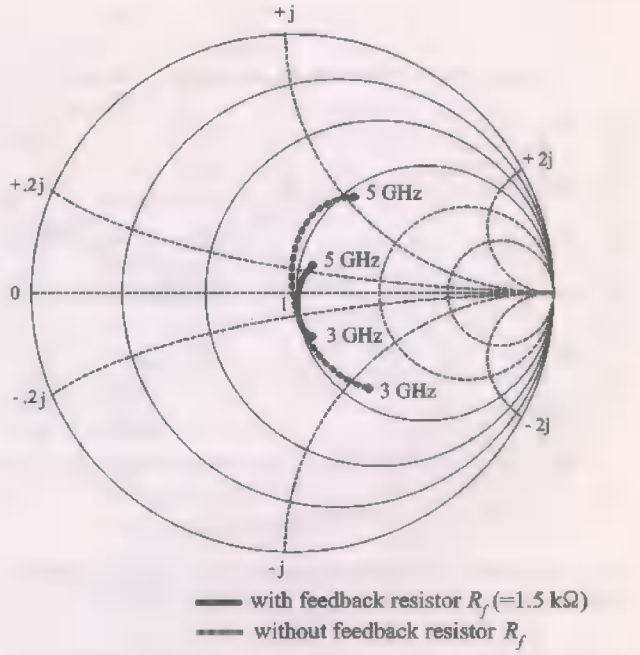


Fig. 3. Simulated S_{11} traces of LNA with or without the feedback resistor for frequencies over 3–5 GHz.

of reactive elements is chosen to resonate at the frequencies of interest such that Z_{in} becomes a real value with $\omega_T L_s$ being equal to R_s. The ω_T represents the cutoff frequency of transistor M_1. The quality factor Q of the series resonating input circuit shown in Fig. 1(b) can be given by [11]

$$Q_{NB} = \frac{1}{(R_S + \omega_T L_S) \cdot \omega_0 \cdot C_{gs}} \tag{1}$$

where ω_0 represents the resonant frequency. With a typical LNA, the Q-factor shown in (1) is generally preferred to be high for high-gain and low-noise performance while dissipating low dc power. Since the fractional -3-dB bandwidth of a typical RLC series resonant circuit is inversely proportional to its Q-factor ($\mathrm{BW}_{-3\mathrm{dB}} = \omega_0/Q_{NB}$), the LNA shown in Fig. 1(a) is unsuitable for wideband application.

Fig. 2(a) shows the proposed wideband LNA topology. In Fig. 2(a), R_f is added as a shunt-feedback element to the conventional cascode narrowband LNA and L_{load} is used as shunt peaking inductor at the output [12]. The capacitor C_f is used for the ac coupling purpose. The source follower, composed of M_3 and M_4, is added for measurement proposes only, and provides wideband output matching. C_1 and C_2 are ac coupling capacitors.

Fig. 2(b) shows the small-signal equivalent circuit for the input part of the proposed wideband LNA. In Fig. 2(b), the resistor $R_{fM}[= R_f/(1 - A_v)]$ represents the Miller equivalent input resistance of R_f, where A_v is the open-loop voltage gain of the LNA. From Fig. 2(a) and (b), the value of R_f can be much larger than that of the conventional resistive shunt-feedback. In the conventional resistive shunt-feedback, the size of R_f is limited as R_{fM} determines the input impedance. However, in the proposed topology, the input impedance is determined by $\omega_T L_s$. Therefore, in Fig. 2(a), one of the key roles of the feedback resistor R_f is to reduce the Q-factor of the resonating narrowband LNA input circuit. The Q-factor of the circuit shown in Fig. 2(b) can be approximately given by

$$Q_{WB} \approx \frac{1}{\left[R_S + \omega_t L_S + \frac{(\omega_0 L_g)^2}{R_{fM}}\right] \cdot \omega_0 \cdot C_{gs}}. \tag{2}$$

From (2), and considering the inversely linear relation between the -3-dB bandwidth and the Q-factor, the narrowband LNA in Fig. 2(a) can be converted into a wideband amplifier by the proper selection of R_f.

For example, to design a wideband amplifier that covers a certain frequency band, the narrowband amplifier will be optimized at the center frequency. Then, the -3-dB bandwidth of the small-signal equivalent input circuit can be set by the proper selection of R_f. Depending on the amount of bandwidth, the required value of R_f can vary and so will the amount of noise contribution by R_f. Fig. 3 shows the simulated S_{11} of the designed UWB amplifier with $R_f(= 1.5\,\mathrm{k}\Omega)$ and compares that of the amplifier without the feedback resistor R_f. As can be seen in Fig. 3, compared to the narrowband case, the addition of R_f gathers the values of passband S_{11} closer to the center of the Smith chart, leading to wideband input matching. The feedback resistor R_f also provides its conventional roles of flattening the gain over a wider bandwidth of frequencies with much smaller noise figure degradation.

III. Amplifier Design and Measurement Results

The proposed topology shown in Fig. 2(a) is applied to a 3.1–5-GHz wideband amplifier based on 0.18-μm CMOS technology. The narrowband LNA is optimized at 4 GHz by the

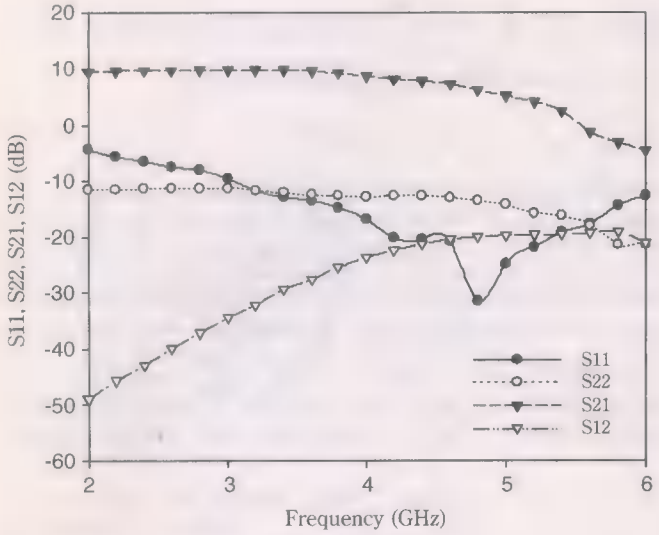


Fig. 4. Measured power gain, input/output return loss, and reverse isolation of the UWB LNA.

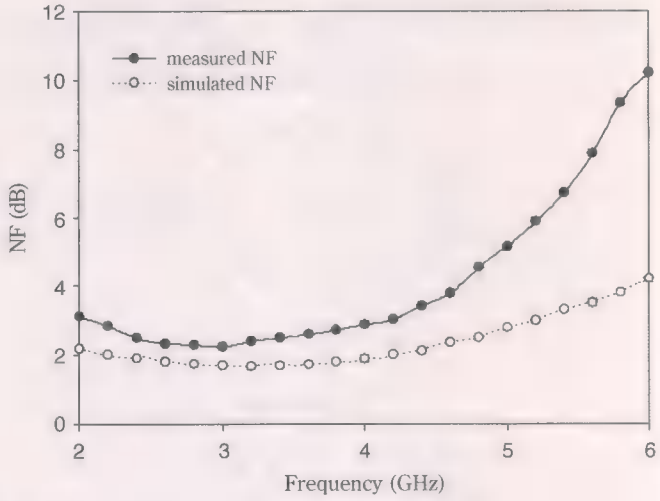


Fig. 5. Measured and simulated NF of the UWB LNA.

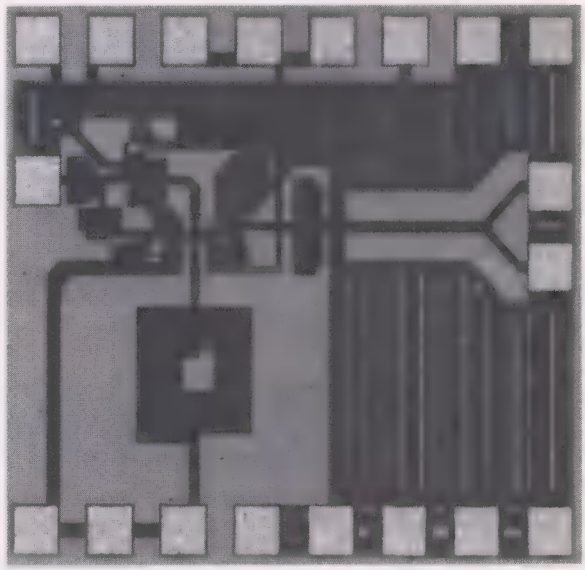


Fig. 6. Microphotograph of the fabricated UWB CMOS LNA. The inductors L_g and L_s are implemented as external components.

proper selection of the values for L_s and L_g. With feedback resistor R_f, the bandwidth extends to cover 3–5 GHz. In Fig. 2(a), the input transistor $M_1(W/L = 320/0.18\ \mu\text{m})$ is biased at 7 mA. The size of the cascode transistor $M_2(240/0.18\ \mu\text{m})$ is decided considering a trade-off between gain (S_{21}) and -3-dB bandwidth. The value of the on-chip spiral inductor L_{load} is 2.4 nH, and its quality factor (Q) is about 9.5 at 5 GHz. The source follower, which consists of $M_3(80/0.18\ \mu\text{m})$ and $M_4(40/0.35\ \mu\text{m})$, consumes 2 mA. Although $R_f = 1.5\ \text{k}\Omega$ is optimal from the simulation results due to the respectable noise performance, the value of R_f is adjusted as 1 kΩ in order to guarantee wideband input matching. In Fig. 2(a), the inductors L_s and L_g are implemented as external components with a value of 0.6 nH and 2.5 nH, respectively. These inductors can be absorbed as a part of the package parasitics, but in this work they are implemented with bond wires due to the chip-on-board (COB) evaluation of the fabricated chip. Other component values are $C_1 = C_f = 2$ pF, $C_2 = 4$ pF, and $R_{\text{load}} = 50\ \Omega$.

For the evaluation, from Fig. 2(a), the dc biasing nodes V_{b1}, V_{b2}, and $V_{DD1} = V_{DD2}$ are biased separately through external voltage sources. Fig. 4 shows the measured S-parameters of the designed UWB amplifier. As can be seen in Fig. 4, the measured input return loss (S_{11}) is higher than 9.0 dB over a 3–5-GHz range. The output return loss (S_{22}) is higher than 11 dB for the same frequency range due to the source follower output stage. The maximum power gain (S_{21}) is +9.8 dB and the -3-dB bandwidth covers 2–4.6 GHz. In Fig. 4, the amplifier shows early power gain roll off near 4.6 GHz compared to the simulated value of 5 GHz. This is caused by the increase in value of the peaking inductance due to the addition of external bonding wires to the supply voltage, which had not been counted properly during the simulation. As can be seen from Fig. 4, the reverse isolation (S_{12}) approaches the 20-dB range due to the feedback network. Considering the reverse isolation provided by the source follower stage, the amount of reverse isolation is worse than expected. Fig. 5 shows both the measured and simulated NF of the implemented amplifier. The measured NF shows a minimum value of 2.3 dB at 3 GHz and stays at less than 3 dB up to 4 GHz, but rises up to 5.2 dB at 5 GHz. Compared to the simulation, the steep increase in NF near 5 GHz is caused by the lower power gain at these frequencies. The discrepancy in NF between the simulation and measurements at the 2–4-GHz range is the result of inaccuracies in the transistor noise model. From the simulation, the feedback resistor R_f degrades the amplifier NF to approximately 0.6 dB. The input referred IP3 is measured as -7 dBm for the two-tone signals of 4 GHz and 4.5 GHz. Fig. 6 shows the microphotograph of the fabricated CMOS UWB LNA with a chip size of 0.9 mm^2. Table I summarizes the measurement results and compares them with previously reported works. In Table I, the indicated amount of power dissipation for this work represents the power dissipated in the cascode topology only.

IV. Conclusion

A new CMOS UWB LNA, applied to the lower band (3.1–5 GHz) UWB system, is presented. The proposed ampli-

TABLE I
COMPARISON OF WIDEBAND CMOS LNA PERFORMANCES: PUBLISHED AND THE PRESENT WORKS

Ref.	$BW_{3\text{-}dB}$ (GHz)	S_{11} (dB)	Gain (dB)	NF* (dB)	IIP3 (dBm)	Power (mW)	Topology	Technology	Year
[3]	0.5 ~ 5.5	< -7	6.5	5.7	-	83.4	Distributed (single-ended)	0.6 μm CMOS	2000
[4]	0.6 ~ 22	< -8	8.1	4.3	-	52	Distributed (single-ended)	0.18 μm CMOS	2003
[5]	0.02 ~ 1.6	< -8	13.7	~1.9	0	35	Feedback (single-ended)	0.25 μm CMOS	2002
[6]	1 ~ 7	< -7.2	13.1	3.3	-4.7	75	Feedback (differential)	0.18 μm CMOS	2003
[8]	2.4 ~ 9.5	< -9.9	9.3	4	-6.7	9**	*LC*-filter based (single-ended)	0.18 μm CMOS	2004
[9]	2 ~ 10	< -10	21	2.5	-5.5	27**	*LC*-filter based (single-ended)	SiGe	2004
This work	**2 ~ 4.6**	**< -9**	**9.8**	**2.3**	**-7**	**12.6****	**Proposed (single-ended)**	**0.18 μm CMOS**	**2004**

* Minimum NF in pass band ** Only core LNA

fier topology adopts the conventional resistive shunt-feedback onto a narrowband LNA topology. In the proposed topology, the wideband characteristics are obtained by utilizing the feedback resistor as a component to reduce the Q-factor of the narrowband amplifier input impedance. The feedback resistor helps to extend the bandwidth of the amplifier as well as the gain flatness, while contributing a small amount in NF degradation. The adoption of the narrowband amplifier allows lower amounts of dc power dissipation. The proposed topology is applied for a 3.1–5-GHz UWB amplifier implementation based on 0.18-μm CMOS technology. The measured results shows more than 9 dB of input return loss, a higher than 11 dB output return loss, a peak gain of 9.8 dB over the -3-dB bandwidth of 2–4.6 GHz, while dissipating 7 mA from a 1.8-V supply. The minimum NF is 2.3 dB at 3 GHz and stays at less than 3 dB up to 4 GHz, but rises up to 5.2 dB at 5 GHz. The proposed LNA shows advantages in overall performance (NF, power gain, power dissipation, chip size, number of external components, etc.), compared to the distributed, conventional shunt-feedback, or filter-based amplifiers that make up other wideband topologies.

REFERENCES

[1] "Multi-band OFDM Physical Layer Proposal," IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), http://grouper.ieee.org/groups/802/15/pub/2003/Jul03/03267r5P802_15_TG3a-Multi-band-OFDM-CFP-Presentation.ppt.

[2] "XtremeSpectrum CFP Presentation," IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), http://grouper.ieee.org/groups/802/15/pub/2003/Jul03/03153r9P802-15_TG3a-XtremeSpectrum-CFP-Presentation.ppt.

[3] B. M. Ballweber, R. Gupta, and D. J. Allstot, "A fully integrated 0.5–5.5-GHz CMOS distributed amplifier," *IEEE Trans. Solid-State Circuits*, vol. 35, no. 2, pp. 231–239, Feb. 2000.

[4] R.-C. Liu, K.-L. Deng, and H. Wang, "A 0.6–22 GHz broadband CMOS distributed amplifier," in *Proc. IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, June 8–10, 2003, pp. 103–106.

[5] F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Noise canceling in wideband CMOS LNA's," in *IEEE ISSCC Dig. Tech. Papers*, vol. 1, Feb. 2002, pp. 406–407.

[6] S. Andersson, C. Svensson, and O. Drugge, "Wideband LNA for a multistandard wireless receiver in 0.18 μm CMOS," in *Proc. ESSCIRC*, Sep. 2003, pp. 655–658.

[7] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York: McGraw-Hill, 2001.

[8] A. Bevilacqua and A. M. Niknejad, "An ultra-wideband CMOS LNA for 3.1 to 10.6 GHz wireless receiver," in *IEEE ISSCC Dig. Tech. Papers*, 2004, pp. 382–383.

[9] A. Ismail and A. Abidi, "A 3 to 10 GHz LNA using a wideband LC-ladder matching network," in *IEEE ISSCC Dig. Tech. Papers*, 2004, pp. 384–385.

[10] T.-K. Nguyen *et al.*, "CMOS low noise amplifier design optimization techniques," *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 5, pp. 1433–1442, May 2004.

[11] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[12] S. S. Mohan *et al.*, "Bandwidth extension in CMOS with optimized on-chip inductors," *IEEE J. Solid-State Circuits*, vol. 35, no. 3, pp. 346–355, Mar. 2000.

CMOS Wideband Amplifiers Using Multiple Inductive-Series Peaking Technique

Chia-Hsin Wu, *Student Member, IEEE*, Chih-Hun Lee, Wei-Sheng Chen, and Shen-Iuan Liu, *Senior Member, IEEE*

***Abstract*—This paper presents the technique of multiple inductive-series peaking to mitigate the deteriorated parasitic capacitance in CMOS technology. Employing multiple inductive-series peaking technique, a 10-Gb/s optical transimpedance amplifier (TIA) has been implemented in a 0.18-μm CMOS process. The 10-Gb/s optical CMOS TIA, which accommodates a PD capacitor of 250 fF, achieves the gain of 61 dBΩ and 3-dB frequency of 7.2 GHz. The noise measurement shows the average noise current of 8.2 pA/$\sqrt{\text{Hz}}$ with power consumption of 70 mW.**

***Index Terms*—Inductive-series peaking, transimpedance amplifier, wideband amplifier.**

I. Introduction

With the rapid proliferation of numerous multimedia networking applications, wideband high-speed telecommunication systems, such as 10-Gb/s optical fiber-link applications, are required. These high-speed front-end circuits [1], [2] are required to be high frequency, low cost, and low power dissipation. However, CMOS devices pose difficult design challenges, such as severe parasitic capacitance, lower transconductance, and noise performance, which mandate circuit innovations to tackle with these issues.

The purpose of this paper is to introduce multiple inductive-series peaking technique to overcome the limitations of CMOS technology. This technique can significantly extend circuit bandwidth without penalty of power consumption. Meanwhile, it can have a relatively flat frequency response similar to *LC*-ladder filters. A 10-Gb/s optical transimpedance amplifier (TIA) has been implemented in 0.18-μm CMOS technology to demonstrate the technique of bandwidth extension.

The design of a TIA should meet stringent constraints, such as gain, bandwidth, noise, and dynamic range. With a typical received power of -15 dBm and a photodiode of responsibility of about 0.75 A/W, TIA must afford more than 1 kΩ (60 dBΩ) transimpedance gain to amplify the weak input current to a detectable signal level for the succeeding stage, such as limiting amplifier [3]. Besides, dynamic range has been a critical issue especially for optical fiber links applications. For low-speed optical interconnects, inverter-configuration TIA has been widely adopted [4]. Nevertheless, for high-speed optical fiber link application, such as more than 2.5 Gb/s, inverter-configuration TIA is seldom used due to its low-speed property. In this paper, the inverter-configuration TIA employing the multiple inductive-series peaking technique has been exploited up to 10–Gb/s in CMOS technology, which also possesses low-power and area-efficient features.

The paper is organized as follows. Section II introduces the proposed multiple inductive-series peaking technique. The circuit designs and schematics are also described in this section. Section III presents experimental results of the TIA. Finally, conclusions are given in Section IV.

Manuscript received May 18, 2004; revised July 23, 2004. This work was supported in part by MediaTek Inc. and the MediaTek Fellowship.

The authors are with the Graduate Institute of Electronics Engineering and Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan 10617, R. O. C. (e-mail: lsi@cc.ee.ntu.edu.tw).

Digital Object Identifier 10.1109/JSSC.2004.840979

II. Multiple Inductive-Series Peaking Technique

The proposed wideband amplifier architecture is shown in Fig. 1(a), where on-chip inductors have been deployed between gain stages. Without employing inductors, amplifier bandwidth is mainly determined by *RC* time constants of every node. In CMOS technology, severe parasitic capacitance deteriorates bandwidth significantly. In the proposed architecture, between gain stages, deployed inductors and parasitic capacitances resemble as a third-order *LC*-ladder filter to perform an impedance transformation network [5], [6].

Considering the inter-stage small-signal model without an inductor in Fig. 1(b), the transfer function can be expressed as

$$\frac{V_{out}}{V_{in}} = \frac{-G_m R_T}{1 + sC_T R_T} \tag{1}$$

where R_T denotes $R_{f1}//R_{f2}$, and C_T represents $C_1 + C_2$. R_{f1}/R_{f2} and C_1/C_2 denote equivalent resistors and capacitors contributed by previous and next stages, respectively. The transfer function of Fig. 1(b) can be derived as shown in (2) at the bottom of the next page. Fig. 2 shows the simulated frequency responses of the first- and third-order filters with different inductances from $0.4L_T$ to $1.6L_T$, where L_T denotes the optimal inductance value, $C_1 = C_2$, and $R_{f1} = R_{f2}$. The simulation results show using smaller inductance can improve bandwidth further but also introduce larger peaking magnitude to deteriorate step response. Employing a proper inductance value L_T with an acceptable overshoot peaking, it can be found that the 3-dB bandwidth of the proposed topology is 2.5 times than that without inserting inductors. The bandwidth-extension effect of proposed technique is more apparent for cascading more stages. Fig. 3 shows the simulated 3-dB bandwidths of wideband amplifiers with different cascading stages, where 3-dB frequencies have been normalized with respect to the 3-dB frequency of first-order *RC* filter. It is shown that the 3-dB bandwidth of the proposed amplifier is 6 times than that of a conventional amplifier, which is a quite large factor. The bandwidth of conventional wideband amplifiers is significantly degraded with cascading more stages. However, that of the proposed wideband

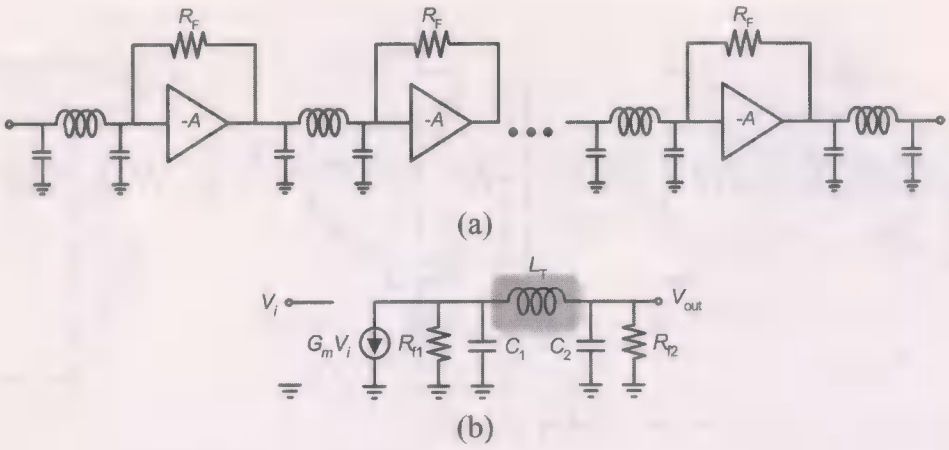


Fig. 1. (a) Proposed wideband amplifier structure. (b) Equivalent inter-stage small-signal model of the proposed amplifier.

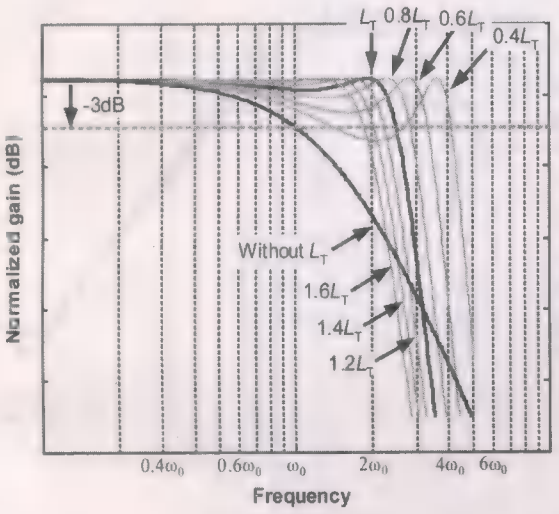


Fig. 2. Comparison between first- and third-order filters with different inductance value.

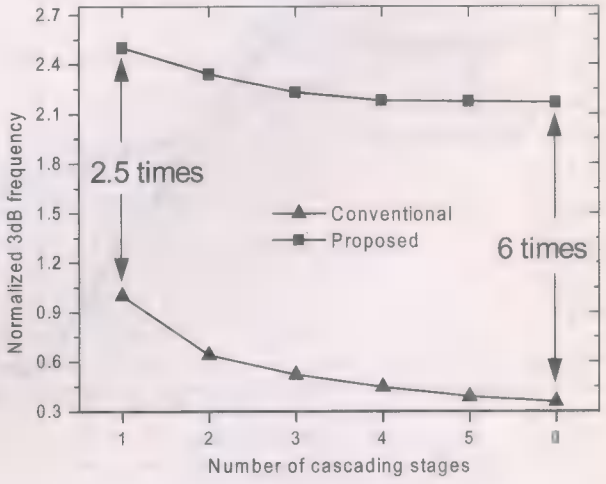


Fig. 3. Simulated 3-dB frequencies versus the number of cascading stages.

amplifier utilizing multiple inductive-series peaking technique is not obviously degraded with cascading more stage, which indicates that the gain and bandwidth trade-off can be ameliorated by the technique.

The proposed TIA is shown in Fig. 4, where on-chip inductors and M-derived half circuits have been employed. Photodiode capacitance, which usually performs the dominant pole, and parasitic capacitances can be absorbed as a part of impedance transformation network by utilizing the multiple inductive-series peaking technique. However, the filter structure performs considerable frequency dependence. If terminated to resistive loads directly, the mismatch will deteriorate the filter significantly. To circumvent this issue, M-derived half circuits, which exhibit more uniform impedance, have been utilized in input and output matching networks [7]. The circuit simulation result is depicted in Fig. 5(a), which shows the 3-dB frequency of conventional 5-stage inverter-configuration TIA is 2.4 GHz, and the 3-dB frequency of the proposed TIA is 7.4 GHz, which is 3 times larger than the conventional one. Considering trade-offs between noise and inter-symbol interference, the bandwidth is commonly determined by 0.7–0.8 times data rate, hence the simulated bandwidth is sufficient for 10-Gb/s optical fiber link application. Fig. 5(b) shows the simulated gains with different inductor series resistance. It is shown that circuit performance is

$$\frac{V_{out}}{V_{in}} = \frac{-G_m R_T}{1 + s\left[C_T R_T + \frac{L_T}{R_{f1}+R_{f2}}\right] + s^2\left[\frac{R_T L_T C_2}{R_{f1}} + \frac{R_T L_T C_1}{R_{f2}}\right] + s^3 C_1 C_2 L_T R_T} \tag{2}$$

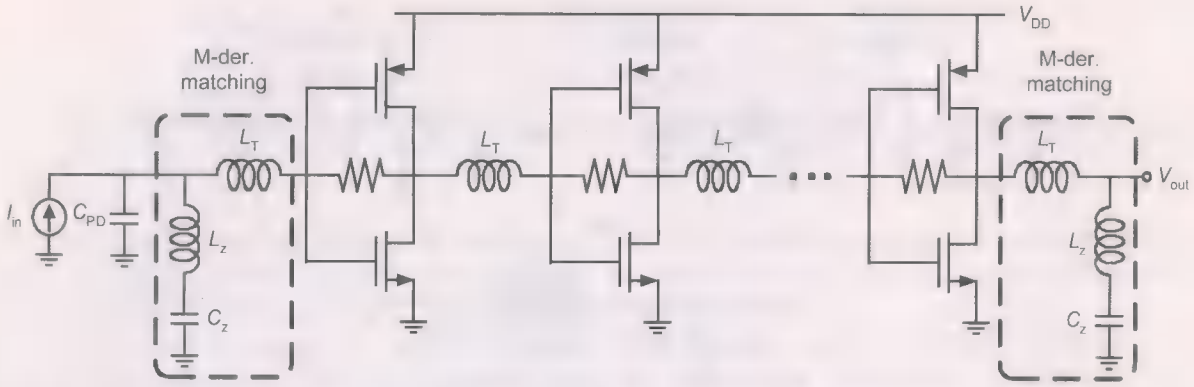


Fig. 4. A 5-stage TIA using proposed multiple inductive-series peaking technique.

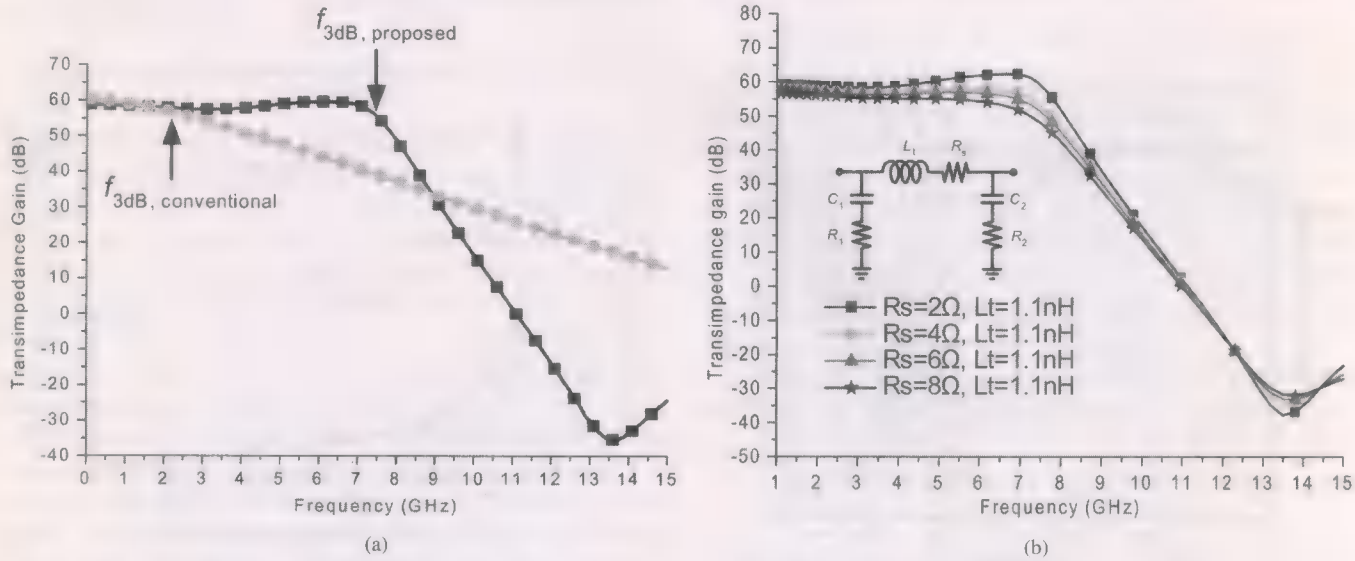


Fig. 5. Simulation results (a) Gains of conventional and proposed TIAs. (b) Proposed TIA's gain versus inductor's series resistance.

insensitive to inductor quality factor. With 50% reduction of inductor quality factor, the gain reduces 2 dB and bandwidth only decreases 3%. Compared to the inductive shunt-peaking technique, which is very sensitive to stray capacitance induced by spiral inductors, the proposed TIA manifests larger bandwidth enhancement and more insensitivity to on-chip inductor quality factor.

Fig. 6. Die photo of the area-efficient TIA.

III. Experimental Results

The proposed TIA has been implemented in 0.18-μm CMOS technology and measured in on-wafer testing. Fig. 6 shows the die photo. To accurately demonstrate the capability of accommodating PD capacitance and load capacitance, two 250-fF MIM capacitors have been integrated on this chip. Ascribed to be insensitive of inductor quality factors, miniature 3-D inductors have been adopted to further minimize die area [8]. The core circuit area is only 0.14 mm^2, which is almost equal to a 5-nH planar inductor.

Fig. 7 shows the measured gain and group delays. The measured gain is 61 dBΩ and 3-dB frequency is 7.2 GHz. Within 3-dB bandwidth, the average group delay is 275 ps with ripple of about 25 ps. Fig. 8 shows the measured average input equivalent noise current density of 8.2 pA/$\sqrt{\text{Hz}}$.

The measured eye diagrams with $2^{31} - 1$ PRBS have been depicted in Fig. 9. The measured output eye diagram is still well open at larger input current of 3.1 mA. Compared to a resistive feedback TIA, the inverter-configuration TIA possesses superior capability to accommodate larger input current. The proposed TIA is well suitable to optical fiber link applications, which needs wide dynamic range requirement.

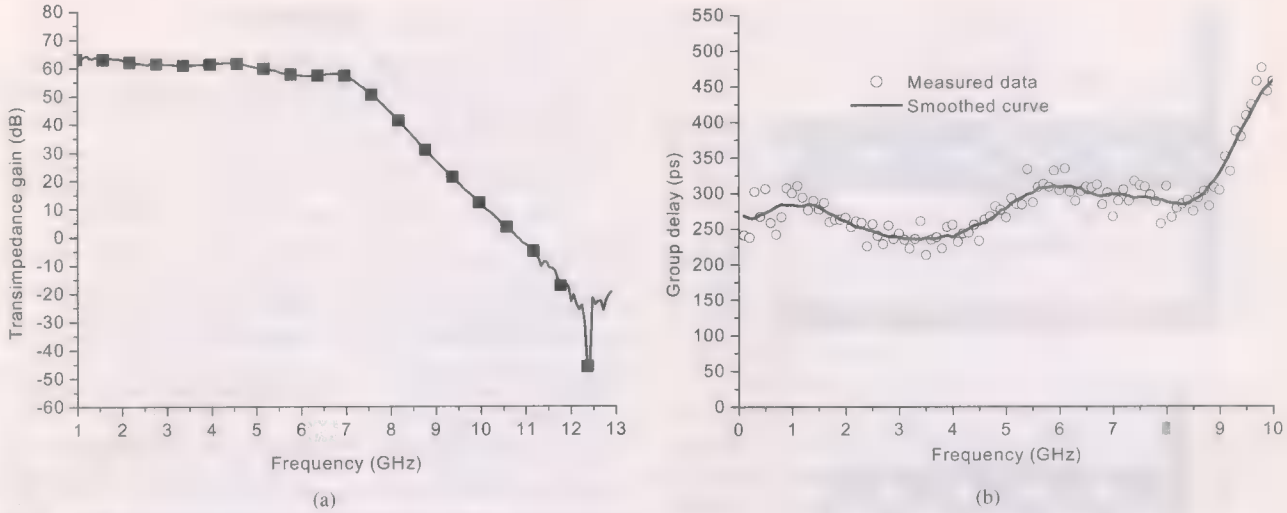


Fig. 7. Measured (a) transimpedance gain and (b) group delays.

TABLE I
SUMMARY OF MEASURED PERFORMANCE AND BRIEF COMPARISON WITH STATE-OF-THE-ART PUBLICATIONS

Reference	This Work	[9]	[10]	[11]
Process	0.18um CMOS	0.18um BiCMOS	0.18um CMOS	0.25um BiCMOS
Supply Voltage	1.8V	2.5V	1.8V	5V
Trans. Gain	1.12 kΩ 61 dBΩ	500 Ω 54 dBΩ	1.3k Ω 62.3 dBΩ	560 Ω 55 dBΩ
-3dB Bandwidth	7.2GHz	9.2GHz	9GHz	9GHz
Speed	10 Gb/s	10 Gb/s	10 Gb/s	10 Gb/s
PD Capacitance	0.25 pF	0.5 pF	0.15 pF	0.15 pF
Sensitivity	10uA (I_{in})	---	---	-17dBm (P_{in})
Input Equivalent Noise	8.2pA/√Hz	< 7pA/√Hz	N/A	9.5pA/√Hz
Power Dissipation	70.2 mW	138 mW	108 mW	140 mW
Chip Area	0.14mm^2	0.64mm^2	N/A	N/A

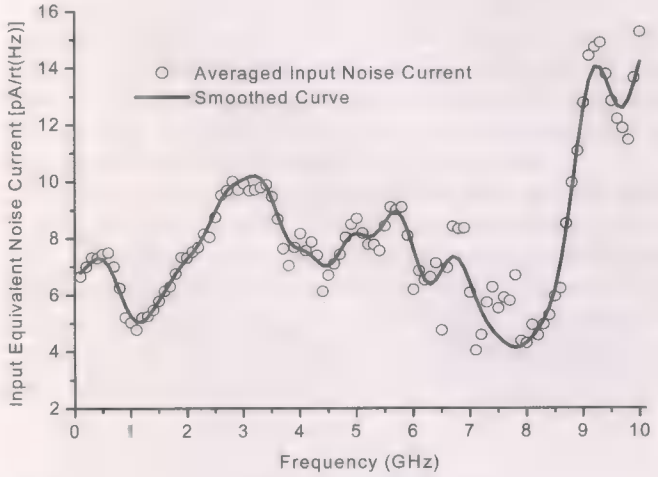
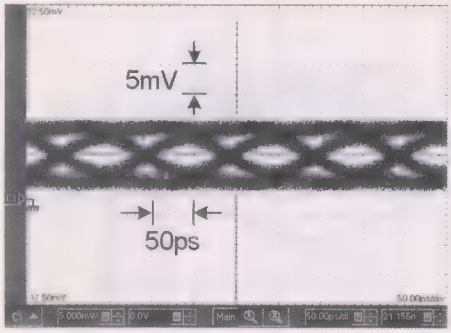


Fig. 8. Measured input equivalent noise current density.

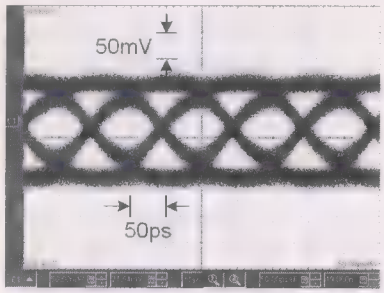
Measured results and the brief comparison with the state-of-the-art 10-Gb/s TIA publications are summarized in Table I. A low-voltage and low-power operation can be achieved by eliminating power-hungry intermediate and output buffers. This fully integrated TIA demonstrates the efficiency of chip area and power consumption, only 0.14 mm^2 and 70.2 mW with a single 1.8-V supply.

IV. CONCLUSION

A bandwidth-extension technique called multiple inductive-series peaking technique has been introduced in this paper. A 10-Gb/s CMOS TIA has been presented to demonstrate the bandwidth-extension technique. Employing the multiple inductive-series peaking technique, the CMOS TIA reported here achieves gain of 61 dBΩ with bandwidth of 7.2 GHz. The measured results demonstrate that the proposed technique of bandwidth extension can improve bandwidth performance significantly. The proposed technique of bandwidth extension is suitable for CMOS devices to achieve wideband and low-power characteristics simultaneously.



(a)



(b)

Fig. 9. Measured eye diagrams. (a) $I_{in} = 10\ \mu$A. (b) $I_{in} = 0.17$ mA with 10 Gb/s $2^{31} - 1$ PRBS.

ACKNOWLEDGMENT

The authors would like to thank CIC for the fabrication of the chip, NDL for measurements, and Prof. H.-W. Tsao for equipment facility.

REFERENCES

[1] J. Cao, A. Momtaz, K. Vakilian, M. M. Green, D. Chung, K.-C. Jen, M. Caresosa, B. Tan, I. Fujimori, and A. Hairapetian, "OC-192 receiver in standard 0.18 μm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2002, pp. 250–251.

[2] S. Galal and B. Razavi, "10 Gb/s limiting amplifier and laser/modulator driver in 0.18 μm CMOS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2003, pp. 188–189.

[3] B. Razavi, *Design of Integrated Circuits for Optical Communications*. New York: McGraw Hill, 2003.

[4] M. Ingels, V. der Plas, J. Crols, and M. Steyaert, "A CMOS 18 THzΩ 240 Mb/s transimpedance amplifier and 155 Mb/s led-driver for low cost optical fiber links," *IEEE J. Solid-State Circuits*, vol. 29, no. 12, pp. 1552–1559, Dec. 1994.

[5] R. Schaumann and M. E. Valkenburg, *Design of Analog Filters*. New York: Oxford Univ. Press, 2001.

[6] S. Galal and B. Razavi, "A 40 Gb/s amplifier and ESD protection circuit in 0.18 μm CMOS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2004, pp. 480–481.

[7] T. T. Y. Wong, *Fundamentals of Distributed Amplification*. Norwood, MA: Artech House, 1993.

[8] C.-C. Tang, C.-H. Wu, and S.-I. Liu, "Miniature 3D inductors in standard CMOS process," *IEEE J. Solid-State Circuits*, vol. 37, no. 4, pp. 471–480, Apr. 2002.

[9] B. Analui and A. Hajimiri, "Multi-pole bandwidth enhancement technique for transimpedance amplifiers," in *ESSCIRC Dig. Tech. Papers*, Sep. 2002, pp. 303–306.

[10] A. K. Peterson, K. Kiziloglu, T. Yoon, F. Williams, and M. R. Sandor Jr, "Front-end CMOS chipset for 10 Gb/s communications," in *IEEE RFIC Dig. Papers*, Jun. 2002, pp. 93–96.

[11] H. H. Kim, S. Chandrasekhar, C. A. Burrus Jr, and J. Bauman, "A Si BiCMOS transimpedance amplifier for 10 Gb/s SONET receiver," *IEEE J. Solid-State Circuits*, vol. 36, no. 5thin, pp. 769–776, May 2001.

60-GHz SOI CMOS Traveling-Wave Amplifier With NF Below 3.8 dB From 0.1 to 40 GHz

Frank Ellinger, *Member, IEEE*

***Abstract*—In this paper, the design and the results of a CMOS traveling-wave amplifier (TWA) optimized for minimum noise figure is presented. Design tradeoffs and optimization guidelines for maximum operation frequency, gain and minimum noise are discussed by means of analytical calculations and simulations. The MMIC is fabricated using digital 90-nm silicon on insulator (SOI) technology and requires a chip area of only 0.3 mm^2. At a supply voltage of 2 V and a supply current of 66 mA, a gain of 9.7 dB$\pm$1.6 dB is measured over a frequency range from 10 to 59 GHz. Toward dc, the gain increases up to 16 dB. The unity gain cutoff frequency is 71 GHz. At 20 and 40 GHz, the circuit has a 1-dB output compression point of 12.5 and 9.5 dBm, respectively. From 0.1 to 40 GHz, a noise figure below 3.8 dB is measured. The results are achieved at source/load impedances of 50 Ω and include the pad parasitics. To the author's knowledge, the TWA has by far the lowest noise figure achieved for a silicon-based amplifier with comparable bandwidth.**

***Index Terms*—CMOS, low-noise amplifier, millimeter-wave frequency, MMIC, SOI, traveling-wave amplifier.**

I. Introduction

OVER the last years, the speed gap between leading-edge III/V and CMOS technologies has been significantly decreased. Recently, a SOI CMOS technology with transit frequency (f_t) of 243 GHz and maximum frequency of oscillation ($f_{\max}$) of 208 GHz has been reported [1]. Compared to conventional bulk technology, the implementation of a thin isolation layer between the active transistor area and the substrate allows a higher substrate resistivity without degrading the threshold properties of the MOSFETs. Consequently, the parasitics of the transistors and the passive devices are reduced thereby increasing their speed and Q factor, respectively.

Analog circuits such as a 26–42-GHz low-noise amplifier [3], a 30–40-GHz mixer [4], a 52–62-GHz oscillator [5] and a 26.5–28.5-GHz frequency doubler [6] have been designed, demonstrating the suitability of SOI CMOS technologies for analog applications at millimeter-wave frequencies.

Wideband amplification is important for many systems such as ultra-wideband (UWB) transceivers, measurement equipment, and optical communication. The excellent bandwidth performance of TWAs is well known [7]. In contrast to cascaded amplifier topologies, the gain of the traveling-wave amplifier (TWA) stages is added instead of multiplied. Thus, TWAs provide a relative low gain. However, due to the incorporation of the parasitic capacitances of the amplifier stages into artificial transmission lines, very high bandwidths can be achieved. Recently, a SOI TWA has been reported yielding a gain of 5 dB up to a very high operation frequency of 91 GHz [8].

In this paper, a TWA is presented, which was optimized for minimum noise and maximum gain up to 40 GHz. The circuit was fabricated on very large scale integration (VLSI) SOI CMOS technology optimized for digital rather than for analog applications. With a noise figure below 3.8 dB from 0.1 to 40 GHz, the presented TWA significantly improves the state-of-the-art noise performance of CMOS wideband amplifiers operating at millimeter-wave frequencies. The result is close to the one achieved with leading-edge III/V technologies. As an example, a TWA on metamorphic HEMT technology has been reported providing a noise figure below 3.7 dB from 5 to 40 GHz [10]. A comparison with other state-of-the-art TWAs is shown in Table I.

II. Modeling

The TWA was fabricated on experimental 90-nm IBM VLSI SOI CMOS technology featuring a metal stack with 8 metals and a substrate resistivity of $13.5\pm5\ \Omega$cm. Detailed information about the technology can be found in [1]–[6].

In Fig. 1, the small-signal and noise model of the n-channel FETs with gate width w_g of 64 μm is shown. It is applied in the HP advanced design system (ADS). The measured and simulated S-parameters and the 50-Ω noise figures are compared in Fig. 2. The device is biased in class-A operation with a drain-source voltage of $V_{ds} = 1$ V, a gate-source voltage of $V_{gs} = 0.5$ V, and a corresponding drain-source current of $I_{ds} = 17$ mA. In this bias point, a f_t of 147 GHz and a $f_{\max}$ of 150 GHz were extracted. The transistors have a threshold voltage of approximately 0.27 V and a drain-source breakdown voltage well above 1 V. At 26 GHz, the FETs have a $\mathrm{NF}_{\min}$ of approximately 1.1 dB [2].

Inductive transmission lines with an inductance per length of approximately 0.7 nH/mm and a loss of around 1.8 dB/mm are used. To minimize the parasitic ground capacitances and to allow high resonance frequencies in the range of 100 GHz, no ground shields are used. For further information about the inductive lines, the reader is referred to [3].

III. Circuit Design

In Fig. 3, the circuit schematics of the designed TWA is shown. The input signal travels down the input line, feeding each amplifier. Undesired reflections are absorbed by the termination resistors R_{abg} and R_{abd}. Given that the phases of the input and output lines are equal, the amplified signals

Manuscript received December 11, 2003; revised August 30, 2004.

The author is with the Swiss Federal Institute of Technology (ETH) Zurich, Electronics Laboratory, Zurich, CH-8092 Switzerland, and also with the IBM/ETH Center for Advanced Silicon Electronics, Zurich, Switzerland (e-mail: ellinger@ife.ee.ethz.ch).

Digital Object Identifier 10.1109/JSSC.2004.840971

TABLE I
STATE-OF-THE-ART TWAs

Technology/f_{max}	Operation BW*	Min. Gain	P_{1dB}	NF	P_{dc}	Chip area	Ref.
III/V based technology							
0.6μm GaAs MESFET/18GHz	12GHz	7dB	n.a.	n.a.	4V×19mA	1.36mm^2	[11]
0.1μm metamorphic HEMT/n.a.	40GHz	14dB	4dBm@20GHz	<3.7dB 5-40GHz	3.5V×143mA	6.3mm^2	[10]
InP HBT/370GHz	80GHz	7.5dB	n.a.	n.a.	250mW	0.84mm^2	[12]
0.1μm InP HEMT/300GHz	112GHz	4dB	n.a.	n.a.	n.a.	2.2mm^2	[13]
Silicon based technology							
0.5μm CMOS/n.a.	5.5GHz	5.3dB	11.8dBm@5GHz	5.5dB@2GHz	3V×27mA	0.79mm^2	[14]
0.18μm CMOS/n.a.	10GHz	8dB	n.a.	n.a.	n.a.	2.34mm^2	[15]
BJT/70GHz	15GHz	8.7dB	n.a.	8.3dB@8GHz	n.a.	7.5mm^2	[16]
0.18μm CMOS/n.a.	22GHz	6.5dB	n.a.	6.1dB@18GHz	1.3V×40mA	1.35mm^2	[17]
SiGe HBT/100 GHz	81GHz	13dB	n.a.	n.a.	5.5V×35mA	2.21mm^2	[18]
0.12μm SOI CMOS/200GHz	91GHz	5dB	n.a.	n.a.	2.6V×35mA	0.82mm^2	[8]
90nm SOI CMOS/160GHz	**59GHz**	**8dB**	**12.5dBm@20GHz**	**<3.8dB 0.1-40GHz**	**2V×66 mA**	**0.3mm^2**	**This work**

*Lower frequency depends on external decoupling capacitors (BW: bandwidth).

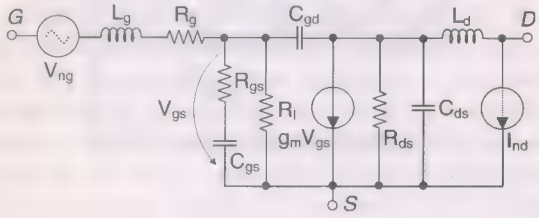


Fig. 1. Small signal and noise model of MOSFET at $V_{gs} = 0.5$ V, $V_{ds} = 1$ V and $I_{ds} = 17$ mA, transconductance $g_m = 82$ mS, drain-source resistance $R_{ds} = 67\ \Omega$, drain inductance $L_d = 35$ pH, gate resistance $R_g = 3\ \Omega$, gate-source resistance $R_{gs} = 20\ \Omega$, gate leakage resistance $R_l = 10$ kΩ, gate-source capacitance $C_{gs} = 60$ fF, gate inductance $L_g = 30$ pH, gate-drain capacitance $C_{gd} = 20$ fF, drain-source capacitance $C_{ds} = 15$ fF, drain noise current source $I_{nd} = 45$ pA, gate noise current source $V_{ng} = 200$ pV.

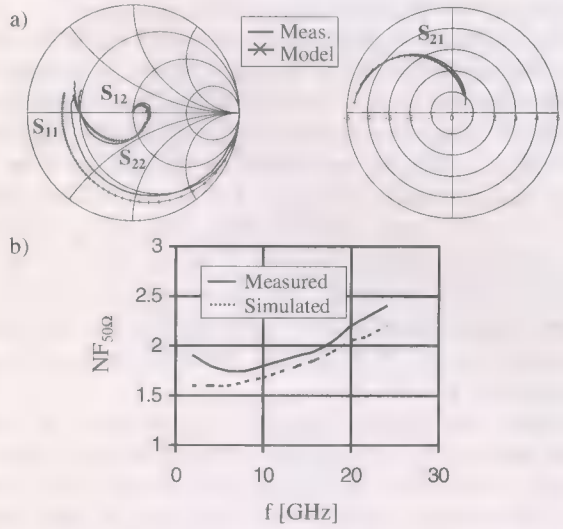


Fig. 2. Comparison between measurements and simulations of MOSFET at $V_{gs} = 0.5$ V, $V_{ds} = 1$ V, $I_{ds} = 17$ mA. (a) S-parameters 2–100 GHz. (b) Noise figure at 50 Ω ($NF_{50\Omega}$).

are constructively added at the output line. This is the case when the values for the inductance L and the capacitance C of the input and output lines are equal. For simplification, it is assumed that the feedback from the input to the output of the amplifiers S_{12} and the parasitics of the inductors can be neglected. Consequently, the capacitance of the distributed line is determined by the input capacitance C_{in} of the amplifier, which typically is larger than the output capacitance. An additional shunt capacitance can be added at the output of the amplifier stages to obtain equal capacitances and phase conditions.

Common-gate and common-drain stages are not well suited for the TWA amplifier stages, since they have resistive rather than capacitive input and output impedances, respectively, thereby causing high line losses. Cascode amplifier stages as illustrated in Fig. 4 were used for the designed TWA, since compared to common-source stages, they provide a significantly higher output impedance with a value above $g_m R_{ds}^2 = 450\ \Omega$, which is approximately 6 times larger than the one of a common-source stage using the same transistor. Due to this high value related to the line impedance of 50 Ω, the output resistance can be neglected for theoretical calculations simplifying the analysis. The resistive output losses of the amplifier can be reduced and the gain can be increased. This is demonstrated in Fig. 5, where the measured power gain of the common-source and the common-gate stages are compared. At 40 GHz, the common-source stage provides a maximum stable gain (MSG) of 10 dB, whereas the cascode stage yields a higher MSG amounting to 17.5 dB.

The characteristic impedance and the 3-dB cutoff frequency of a distributed line section can be approximated by

$$Z_0 = \sqrt{\frac{L}{C_{in}}} \tag{1}$$

and

$$f_c = \frac{1}{\pi\sqrt{LC_{in}}} \tag{2}$$

with L as the line inductance. The choice of w_g is a tradeoff between desired g_m and corresponding power gain per stage on one side, and maximum f_c on the other side. We can determine the maximum C_{in} and the associated w_g for a desired f_c. The design goal of this work was to achieve an operation frequency of at least 40 GHz. To ensure that the f_c of the transmission line sections is well above this frequency, we chose a f_c of 70 GHz. With a Z_0 of 50 Ω, we obtain a w_g of 64 μm and a L of 225 pH.

The power gain of TWAs is limited by the gate line, drain line, and inductive line losses. As discussed before, the losses of the inductive lines are relatively small. It has been shown that the

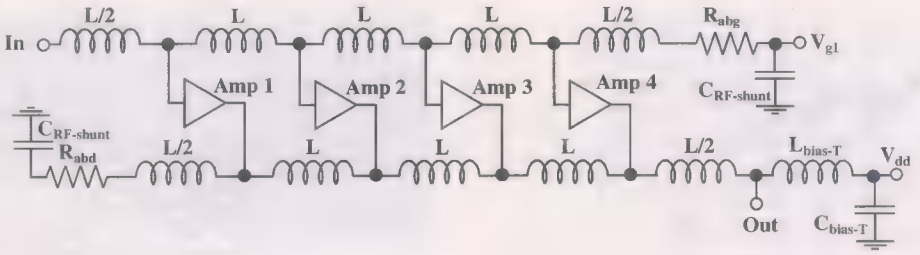


Fig. 3. Circuit schematics of TWA with four stages, $L = 170$ pH, $R_{\rm abg} = 75\ \Omega$, $R_{\rm abd} = 50\ \Omega$, $C_{\rm RF\ shunt} = 25$ pF, $V_{g1} = 0.5$ V, $V_{dd} = 2$ V.

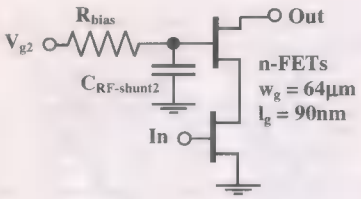


Fig. 4. Simplified circuit schematics of cascode amplifier stage, $C_{\rm RF\ shunt2} = 5$ pF, $R_{\rm bias} = 6\ {\rm k}\Omega$, $V_{g2} = 1.5$ V, V_{ds} of each FET ≈ 1 V.

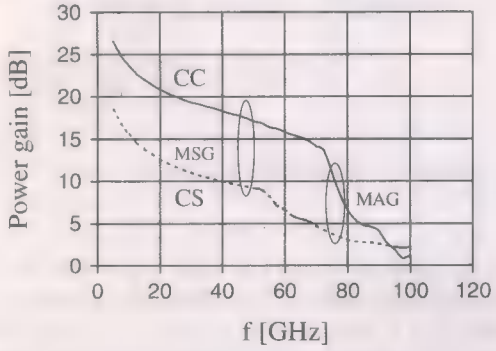


Fig. 5. Measured stable gain (MSG) and maximum available gain (MAG) of common source (CS) and cascode (CC) amplifier stages.

losses are mainly determined by the gate line losses [19]. This is especially the case for TWAs using cascode amplifier stages with high output resistance. If the losses of the drain line and the inductors are neglected, the small-signal power gain can be approximated by

$$G = G_0(1 - nA_g)^2 \tag{3}$$

with the low-frequency gain

$$G_0 = \left(\frac{n \cdot g_m \cdot Z_0}{2}\right)^2 \tag{4}$$

and the gate loss factor

$$A_g = \frac{1}{4} R_g \omega^2 C_{in}^2 Z_0. \tag{5}$$

For derivations and explanations of (3)–(5), the reader is referred to [19]. The third term of (5) from [19] was neglected since its impact is small compared to the second term. Furthermore, we have substituted the factor $a_g l_g/2$ from [19] by A_g. By means of (3)–(5) we can show that for a given frequency, maximum gain is achieved for a number of stages of

$$n_{oG} = \frac{1}{2A_g}. \tag{6}$$

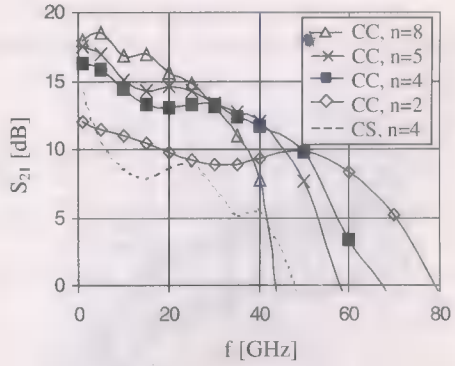


Fig. 6. Simulated gain using cascode (CC) and common source (CS) amplifier stages with different number of stages n.

With the given device parameters and an operation frequency of 40 GHz, where according to the design goal optimum performance should be reached, we obtain $A_g = 0.0895$, $n_{oG} = 5.6$, $G_0 = 21$ dB, and $G = 15$ dB. These calculations are appropriate for first considerations and optimizations. Due to the additional losses generated by the drain line and the inductors, the total line losses will be slightly higher than assumed. Thus, in reality, the values of n_{oG} and G are upper limits.

Furthermore, ADS simulations using the more precise model presented in Fig. 1, and lumped equivalent circuits for the passive devices [3], [4], were performed.

In Fig. 6, the simulated gain versus frequency and number of cascode stages is shown. For a frequency up to 40 GHz, the simulation predict a n_{oG} of approximately 5, which is in good agreement with the theoretical calculations. Due to a more accurate consideration of the parasitics, the simulated power gain of 12 dB at 40 GHz is 3 dB lower than the calculated one. The results of a TWA with common-source stages is also included for comparison verifying the superior properties of the cascode circuit.

The performance of the circuit is influenced by the inductors. In Fig. 7, the simulated gain versus frequency is illustrated for different inductor values. All relevant parasitics are considered for the scalable inductor model. There are the following effects: with increasing inductor value, the capacitive parasitics of the FETs can be compensated improving the gain. However, an increasing inductor value has two drawbacks. First, the series resistance of the inductor becomes large. Furthermore, above an associated resonance frequency, the gain drops significantly. Both effects degrade the maximum gain cutoff frequency. Therefore, an optimum inductor value has to be chosen. According to Fig. 7, a value of $L = 170$ nH is well suited for

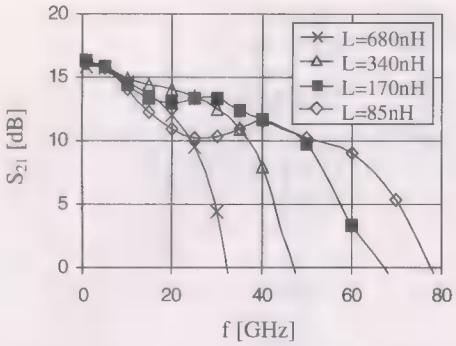


Fig. 7. Simulated gain for different inductor values; parasitics are considered and scaled.

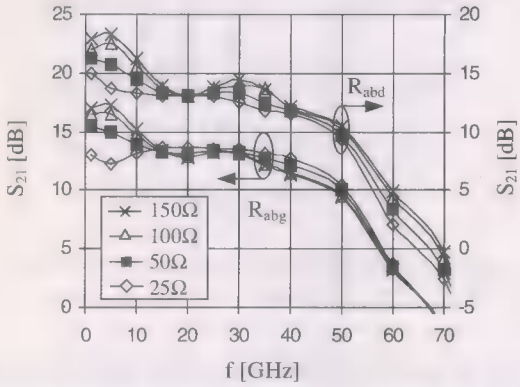


Fig. 8. Simulated gain for different input (R_{abg}) and output (R_{abd}) line termination resistors, bias of each FET.

optimum gain performance up to 40 GHz. This value is slightly lower than the one found by the idealized calculations.

Furthermore, the line termination resistors have a significant impact as depicted in Fig. 8. The gain toward low frequencies decreases with falling resistor values. Thus, the gain flatness and 3-dB bandwidth can be improved. However, we will see later that an decreased input termination resistor degrades the noise performance. Consequently, a high R_{abg} together with a low R_{abd} is advantageous concerning an optimum tradeoff between 3-dB bandwidth and noise.

With the device-dependent drain and gate noise coefficients γ and δ, respectively, the noise figure of FET TWAs can be approximated by [20]

$$F = 1 + \frac{4\gamma}{n \cdot g_m \cdot Z_0} + \frac{n \cdot \omega^2 \cdot C_{in}^2 \cdot Z_0 \cdot \delta}{3g_m}. \tag{7}$$

The second term describes the drain noise, which is dominant at low frequencies, whereas the third term represents the frequency-dependent gate noise determining the high-frequency performance. Typically, values of $2/3 < \gamma < 1$ and $\delta = 4/3$ are reported for long-channel devices [21]. Due to hot electron effects, significantly higher drain noise currents and γ coefficients are expected for short-channel devices as used in this work. By fitting of the measured noise figure, we obtain values of $\gamma = 2.2$

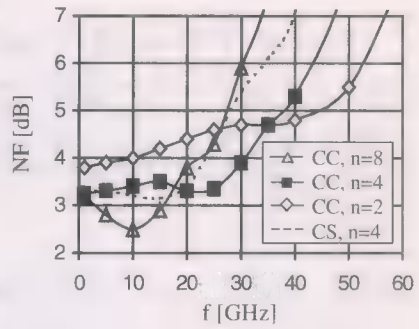


Fig. 9. Simulated noise figure of different TWAs with cascode and common source amplifier stages, n: number of stages.

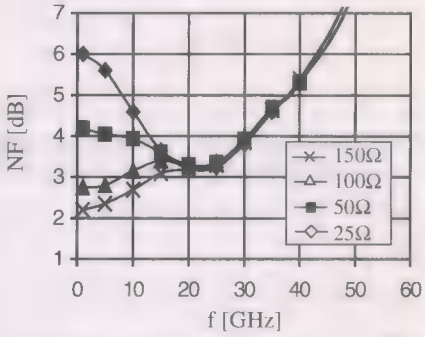


Fig. 10. Simulated noise figure of four-stage cascode TWA with different gate line terminations.

and $\delta = 1.5$, which is in very good agreement with the data extracted for a single transistor [2]. From (7), a minimum noise figure of

$$F_{\min} = 1 + \frac{2\omega C_{in}}{g_m}\sqrt{\frac{4\gamma\delta}{3}} \tag{8}$$

can be derived for a number of stages of

$$n_{oF} = \frac{2}{\omega C_{in} Z_0}\sqrt{\frac{3\gamma}{\delta}}. \tag{9}$$

At an operation frequency of 40 GHz, we obtain $n_{oF} = 3.7$ and $F_{\min} = 3.3$ dB.

For comparison, noise simulations were performed in ADS. As depicted in Fig. 9, up to 40 GHz, good noise performance is achieved for a n_{oF} of approximately 4 verifying the theoretical results. Furthermore, the simulations show that the best low-frequency noise performance is achieved at high n_{oF}, whereas for high frequencies, the lowest noise figures are reached for low values of n_{oF}. In accordance to (7), this is attributed to the fact that the drain noise is inversely proportional to n_{oF}, whereas the gate noise is proportional to n_{oF}.

At low frequencies, a TWA behaves as a single transistor with all amplifier stages connected in parallel. Furthermore, the input and output are directly terminated by the absorption resistors. As clearly shown in Fig. 10, the gate line termination resistor R_{abg} significantly increases the noise toward low frequencies. Thus, the low-frequency noise performance can be improved by increasing the input termination resistor. Unfortunately, this decreases the input return loss. A nominal value of $R_{abg} = 75\,\Omega$

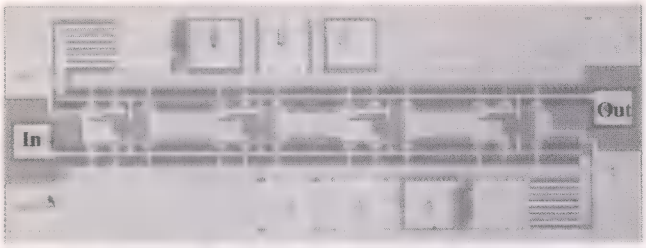


Fig. 11. Photograph of compact TWA MMIC with chip size of 0.89 mm × 0.33 mm.

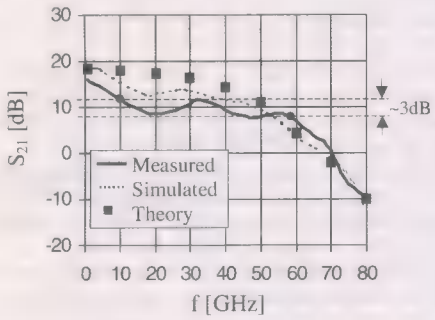


Fig. 12. Measured, simulated, and calculated gain.

was chosen since this provides a reasonable tradeoff between enhanced noise performance and acceptable input return loss.

Up to 40 GHz, the calculated values for n_{oF} and n_{oG} are close together. A value of $n = 4$ was used for the final realization of the TWA. The nominal value of R_{abd} is 50 Ω.

A photograph of the compact TWA MMIC with overall chip size of 0.89 mm × 0.33 mm is shown in Fig. 11. To the author's knowledge, this is the smallest chip size of a TWA reported to date. In mass fabrication, the small chip size scales down the costs.

IV. Results

All measurements were performed on wafer, at source and load impedances of 50 Ω and include the parasitics of the signal pads. The power consumption is $V_{dd} = 2$ V and $I_{dd} = 66$ mA. As for the device characterization, S-parameters were measured using an HP 8510XF network analyzer. The noise figure setup consists of an HP 8970B noise figure meter, an HP 8971C test set extension and a external mixer allowing measurements up to 40 GHz. An HP 436A power meter was used for determination of the compression point.

The measured wafer was based on experimental hardware that showed process variations. The deviation of +60% for the termination resistors and the corresponding impact on the circuit characteristics are significant and were considered in the following simulations.

With a Rollet's factor well above 1, the circuit is unconditionally stable. In Fig. 12, the measured, simulated and calculated gain is shown. A gain of 9.7 dB±1.6 dB was measured from 10 to 59 GHz. Toward dc, the gain increases up to 16 dB. The gain cutoff frequency is 71 GHz.

The measured, simulated and calculated noise figure of the circuit is shown in Fig. 13. Toward dc and between 23 and

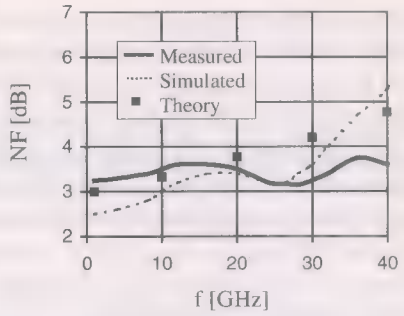


Fig. 13. Measured, simulated and calculated noise figure.

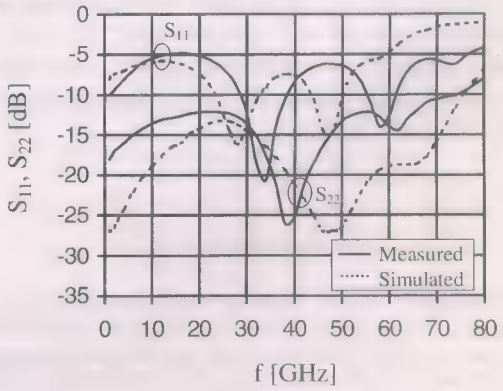


Fig. 14. Measured and simulated return losses.

29 GHz, the noise figure is approximately 3.2 dB. Up to 40 GHz, the noise figure is below 3.8 dB. To the author's knowledge, these are the best results achieved for a silicon-based wideband amplifier operating up to millimeter-wave frequencies. Unfortunately, with our current measurement equipment, it is not possible to characterize the noise figure at higher frequencies.

In Fig. 14, the measured and simulated return losses are shown. From dc to 60 GHz, the measured input and output return losses are higher than 5 and 12 dB. Higher return losses are expected for circuits from more nominal wafers.

At 0.1, 20, and 40 GHz, the measured 1-dB output compression points are 13.3, 12.5, and 9.5 dBm, corresponding to power added efficiencies of 16%, 13.5%, and 6.7%, respectively. The TWA was optimized as a low-noise amplifier. However, due the good large-signal performances, the circuit can also be used as a medium-power amplifier. The output power should be sufficient for many short-range WLAN systems.

V. Conclusion

The design and results of a low-noise CMOS TWA has been presented. Design tradeoffs and optimization guidelines for maximum operation frequency, gain, output power, and minimum noise have been discussed by means of analytical calculations and simulations.

The circuit has been fabricated using 90-nm SOI technology and requires a chip area of only 0.3 mm^2, which to the author's knowledge is the smallest size reported for a TWA. The used technology is optimized for digital VLSI applications rather

than for analog applications. Despite the restrictions of this technology for analog circuits, excellent results have been achieved. From 0.1 to 59 GHz, the circuit has a gain above 8 dB. A very low noise figure of below 3.8 dB has been measured from 0.1 dB to 40 GHz. The author believes that this is best noise performance demonstrated for a silicon-based amplifier with comparable bandwidth. The achieved result is close to the one reported using leading-edge III/V technology. With a 1-dB output compression point of 13.3 to 9.5 dBm from 0.1 to 40 GHz, the circuit is also suited as a medium-power amplifier.

Together with other works, this paper clearly shows the excellent suitability of VLSI SOI CMOS technology for analog circuits at millimeter-wave frequencies, which not long ago were the exclusive domain of III/V technologies. This may lead to new market perspectives in areas such as WLAN, measurement equipment, and radar systems, since in the future, high data rates could be achieved at low costs.

REFERENCES

[1] N. Zamdmer, J. Kim, R. Trzcinski, J.-O. Plochart, S. Narasimha, M. Khare, L. Wagner, and S. Chaloux, "A 243-GHz F_t and 208-GHz F_{max}, 90-nm SOI CMOS SoC technology with low-power millimeter-wave digital and RF circuit capability," in *Symp. VLSI Technology Dig. Tech. Papers*, 2004, pp. 98–99.

[2] F. Ellinger, M. Schmatz, and H. Jäckel, "Noise investigations of 90 nm VLSI CMOS technologies for analog integrated circuits at millimeter wave frequencies," in *SPIE Conf. Noise and Fluctuations*, May 2004, pp. 131–138.

[3] F. Ellinger, "26–42 GHz low noise amplifier MMIC fabricated on digital SOI CMOS technology," *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 522–528, Mar. 2004.

[4] F. Ellinger, L. C. Rodoni, G. Sialm, C. Kromer, G. von Büren, M. Schmatz, C. Menolfi, T. Toifl, T. Morf, M. Kossel, and H. Jäckel, "30–40 GHz drain pumped passive down mixer MMIC fabricated on digital SOI CMOS technology," *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 5, pp. 1382–1391, May 2004.

[5] F. Ellinger, T. Morf, G. von Büren, C. Kromer, G. Sialm, L. Rodoni, M. Schmatz, and H. Jäckel, "60 GHz VCO with high tuning range fabricated on VLSI SOI CMOS technology," in *IEEE MTT-S Microwave Symp. Dig.*, Jun. 2004, pp. 1329–1332.

[6] F. Ellinger and H. Jäckel, "Ultra compact SOI CMOS frequency doubler MMIC for low power applications at 26.5–28.5 GHz," *IEEE Microwave Compon. Lett.*, vol. 14, no. 2, pp. 53–55, Feb. 2004.

[7] A. Hajimiri, "Distributed integrated circuits: An alternative approach to high-frequency design," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 168–173, Feb. 2002.

[8] J.-O. Plouchart, J. Kim, N. Zamdmer, L.-H. Lu, M. Sherony, Y. Tan, R. Groves, R. Trzcinski, M. Talbi, A. Ray, and L. Wagner, "A 4–91 GHz distributed amplifier in a standard 0.12 μm SOI CMOS microprocessor technology," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2003, pp. 159–162.

[9] F. Ellinger, M. Kossel, M. Huber, M. Schmatz, C. Kromer, G. Sialm, D. Barras, L. Rodoni, G. von Büren, and H. Jäckel, "High-Q inductors on digital VLSI CMOS substrate for analog RF applications," in *IEEE Int. Microwave and Optoelectronics Conf.*, Sep. 2003, pp. 869–872.

[10] R. E. Leoni III, S. J. Lichwala, J. G. Hunt, C. S. Whelan, P. F. Marsh, W. E. Hoke, and T. E. Kazior, "A DC-45 GHz metamorphic HEMT traveling wave amplifier," in *Proc. IEEE GaAs Symp.*, 2001, pp. 133–136.

[11] A. Orzati and W. Bächtold, "Monolithically integrated traveling-wave amplifier for low-cost broadband optical receiver," in *Workshop on Compound Semiconductor Devices and Integrated Circuits Europe*, May 2001, pp. 9–10.

[12] B. Agarwal, Q. Lee, D. Mensa, R. Pullela, J. Guthrie, and M. J. W. Rodwell, "80-GHz distributed amplifier with transferred substrate heterojunction bipolar transistors," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 12, pp. 2302–2307, Dec. 1998.

[13] B. Agarwal, A. E. Schmitz, J. J. Brown, M. Matloubian, M. G. Case, M. Le, M. Lui, and M. J. W. Rodwell, "112-GHz, 157 GHz, and 180-GHz InP HEMT traveling-wave amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 12, pp. 2553–2559, Dec. 1998.

[14] B. Ballweber, R. Gupta, and D. Allstot, "A fully integrated 0.5–5.5 GHz CMOS distributed amplifier," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 231–239, Feb. 2000.

[15] B. M. Frank, P. Freundorfer, and Y. M. M. Antar, "Performance of 1–10-GHz traveling wave amplifiers in 0.18 μm CMOS," *IEEE Microwave Compon. Lett.*, vol. 12, no. 9, pp. 327–329, Sep. 2002.

[16] M. Kawashima, H. Hazashi, T. Nakagawa, and K. Araki, "A low-noise distributed amplifier using cascode-connected BJT's terminal circuit," in *Proc. Asia Pacific Microwave Conf.*, Dec. 2001, pp. 21–24.

[17] R.-C. Liu, K.-L. Deng, and H. Wang, "A 0.6–22-GHz broadband CMOS distributed amplifier," in *Proc. IEEE Radio Frequency Integrated Symp.*, Philadelphia, PA, June 2003, pp. 103–106.

[18] O. Wohlgemuth, P. Paschke, and Y. Baeyens, "SiGe broadband amplifiers with up to 80 GHz bandwidth for optical applications at 43 Gbit/s and beyond," in *Proc. Eur. Microwave Conf.*, Munich, Germany, Sep. 2003, pp. 1087–1090.

[19] Y. Ayasli, R. L. Mozzi, J. L. Vorhaus, L. D. Reynolds, and R. A. Pucel, "A monolithic GaAs 1–13-GHz traveling-wave amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-30, no. 7, pp. 976–981, Jul. 1982.

[20] C. S. Aitchison, "The intrinsic noise figure of the MESFET distributed amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-33, no. 6, pp. 460–466, Jun. 1985.

[21] P. J. Sullivan, B. A. Xavier, and W. H. Ku, "An integrated CMOS distributed amplifier utilizing packaging inductances," *IEEE Trans. Microwave Theory Tech.*, vol. 45, no. 10, pp. 1969–1976, Oct. 1997.

[22] D. Ham and A. Hajimiri, "Concepts and methods in optimization of integrated LC VCOs," *IEEE J. Solid-State Circuits*, vol. 36, no. 7, pp. 896–909, Jun. 2001.

[23] A. A. Abidi, "High-frequency noise measurements on FET's with small dimensions," *IEEE Trans. Electron Devices*, vol. ED-33, no. 11, pp. 1801–1805, Nov. 1986.

Correspondence

Addition to "A Wideband 2.4-GHz Delta-Sigma Fractional-N PLL With 1-Mb/s In-Loop Modulation"

Sudhakar Pamarti, *Member, IEEE*, Lars Jansson, *Member, IEEE*, and Ian Galton, *Member, IEEE*

A technique was presented in [1] that is similar to that presented in the above paper [2]. It was published shortly before the above paper went to press, and therefore should have been included as a reference in the above paper.

References

[1] I. Bietti, E. Ternporitil, G. Albasini, and R. Castello, "An UMTS sigma delta fractional synthesizer with 200 kHz bandwidth and −128 dBc/Hz at 1 MHz using spurs compensation and linearization techniques," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2003, pp. 463–466.

[2] S. Pamarti, L. Jansson, and I. Galton, "A wideband 2.4-GHz delta-sigma fractional-N PLL with 1-Mb/s in-loop modulation," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 49–62, Jan. 2004.

Manuscript received July 20, 2004.

S. Pamarti is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: spamarti@gmail.com).

L. Jansson is with Silicon Wave, Inc., San Diego, CA 92122 USA.

I. Galton is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92092 USA (e-mail: galton@ece.ucsd.edu).

Digital Object Identifier 10.1109/JSSC.2004.842370

Correction to "A 40-Gb/s Clock and Data Recovery Circuit in 0.18-μm CMOS Technology"

Jri Lee, *Member, IEEE*, and Behzad Razavi, *Fellow, IEEE*

The first author of [1] has indicated that the topologies shown in Fig. 4 of the above paper [2] are the same as those described in [1], [3], and [4]. He has also stated that the means of detection of the direction of the wave described on page 2184 of [2] is the same as that in [4]. We regret the unintentional omission of these references.

References

[1] J. Wood, T. C. Edwards, and S. Lipa, "Rotary traveling-wave oscillator arrays: A new clock technology," *IEEE J. Solid-State Circuits*, vol. 36, no. 11, pp. 1654–1665, Nov. 2001.

[2] J. Lee and B. Razavi, "A 40-Gb/s clock and data recovery circuit in 0.18-μm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2181–2190, Dec. 2003.

[3] J. Wood, " Electronic circuitry," U.S. Patent 6,556,089 B2, Apr. 29, 2003.

[4] ——, International Patent Application Number PCT/GB01/02069, May 11, 2001.

Manuscript received December 23, 2004.

J. Lee is with the National Taiwan University, Taipei, Taiwan 106, R.O.C.

B. Razavi is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: razavi@ee.ucla.edu).

Digital Object Identifier 10.1109/JSSC.2004.842373

Patent Abstracts

The Patent Abstracts and References cited are intended to provide the minimum information necessary for determining interest. The full text and images can be obtained from the U.S. Patent Office at http://www.uspto.gov.

6,760,238 July 6, 2004

Apparatus and Method for DC/DC Converter Having High Speed and Accuracy

Inventor: Charych; Arthur (Setauket, NY)
Assignee: BC Systems, Inc (Setauket, NY)
Filed: October 24, 2002.

Current U.S. Class : **363/97**; 363/21.11; 363/21.18
Intern'l Class : H02M 003/24
Field of Search : 363/21.05, 21.1, 21.11, 21.13, 21.18, 97, 131, 322/282, 283, 284

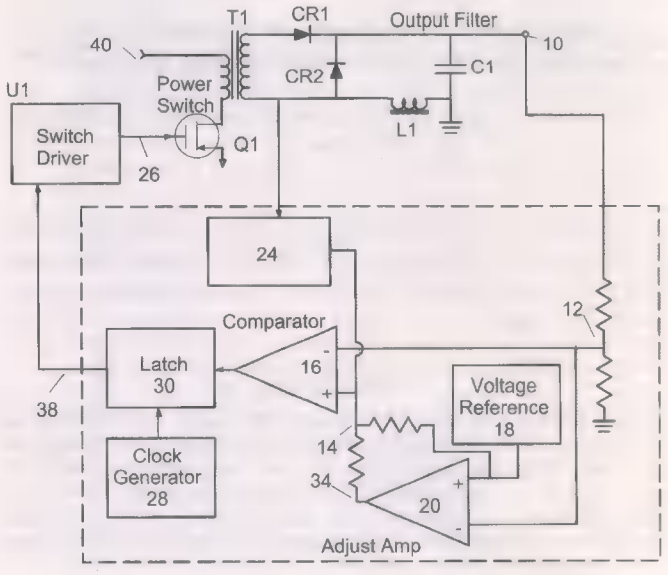
References Cited

U.S. Patent Documents

5260861	Nov., 1993	Wert	363/25.
5770940	Jun., 1998	Goder	323/282.
5912552	Jun., 1999	Tateishi	323/285.
6169680	Jan., 2001	Matsui *et al.*	363/97.
6288524	Sep., 2001	Tsujimoto	323/285.
6304066	Oct., 2001	Wilcox *et al.*	323/282.
6396725	May, 2002	Jacobs *et al.*	363/131.
6434025	Aug., 2002	Shirai *et al.*	363/21.

Abstract*—**A system and method for DC/DC conversion are provided in which a high accuracy ***digital pulse width modulator controller ***circuit*** controls a power switch to obtain a desired DC output. The control ***circuit*** amplifies the difference of a DC output sample in relation to voltage reference. The amplified difference is then compared with a portion of the DC output. The compared result is used for controlling the power switch. A ripple coming from the DC output side is overlaid upon either one of the inputs to the comparator depending upon the polarity of the ripple signal.

Digital Object Identifier 10.1109/JSSC.2004.843081



6,760,266 July 6, 2004

Sense Amplifier and Method for Performing a Read Operation in a MRAM

Inventors: Garni; Bradley J. (Austin, TX), Deherrera; Mark F. (Tempe, AZ), Durlam; Mark A. (Chandler, AZ), Engel; Bradley N. (Chandler, AZ), Andre; Thomas W. (Austin, TX), Nahas; Joseph J. (Austin, TX), and Subramanian; Chitra K. (Austin, TX).
Assignee: Freescale Semiconductor, Inc. (Austin, TX)
Filed: June 28, 2002.

Current U.S. Class : **365/209**; 365/213
Intern'l Class : G11C 007/02
Field of Search : 365/209, 213, 158, 210, 205, 207, 208, 230.07

References Cited

U.S. Patent Documents

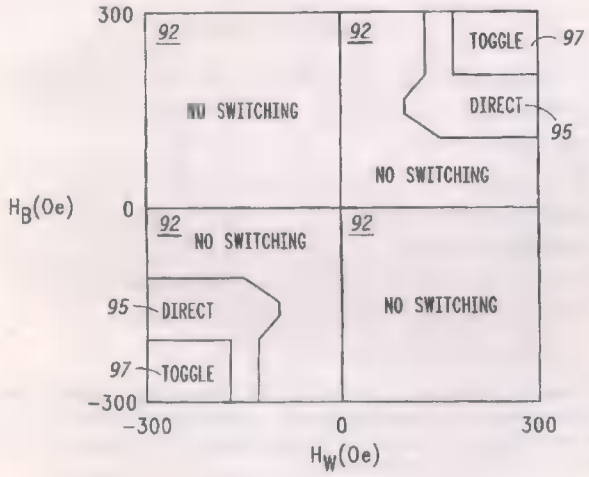
4763026	Aug., 1988	Tsen *et al.*	327/56.
5309393	May, 1994	Sakata *et al.*	365/189.
6188615	Feb., 2001	Perner *et al.*	
6191989	Feb., 2001	Luke *et al.*	
6205073	Mar., 2001	Naji.	
6256247	Jul., 2001	Perner.	
6379978	Apr., 2002	Goebel *et al.*	
6392853	May, 2002	Li *et al.*	
6392924	May, 2002	Lui *et al.*	

Other References

Ranmuthu *et al.*, "A 512K-Bit Magneto Resistive Memory with Switched Capacitor Self-Referencing Sensing," IEEE Transactions on Circuits and Systems-11 : Analog and Digital Signal Processing, 1992, vol. 39, no. 8, pp. 585-587.

Ranmuthu *et al.*, "A Sensing Scheme for Giant Magneto-Resistive Memories," Magnetics Conference, Digest of International Pages. Intermag '93, 1993.

Abstract*—**A sense amplifier (1300, 1500) is provided for sensing the state of ■ toggling type magnetoresistive random access memory (MRAM) cell without using ■ reference. The sense amplifier (1300, 1500) employs a sample-and-hold ***circuit (1336, 1508) combined with ■ current-to-voltage converter (1301, 1501), gain ***circuit*** (1303), and cross-coupled latch (1305, 1503) to sense the state of a bit. The sense amplifier (1300, 1500), first senses and holds a first state of the cell. The cell is toggled to ■ second state. Then, the sense amplifier (1300, 1500) compares the first state to the second state to determine the first state of a toggling type memory cell.



6,762,633 July 13, 2004

Delay Locked Loop Circuit With Improved Jitter Performance

Inventor: Lee; Seong Hoon (Kyoungki-do, KR)
Assignee: Hynix Semiconductor Inc. (Kyoungki-do, KR)
Filed: December 10, 2002.

Foreign Application Priority Data

Dec. 21, 2001[KR] 10-2001-0082675

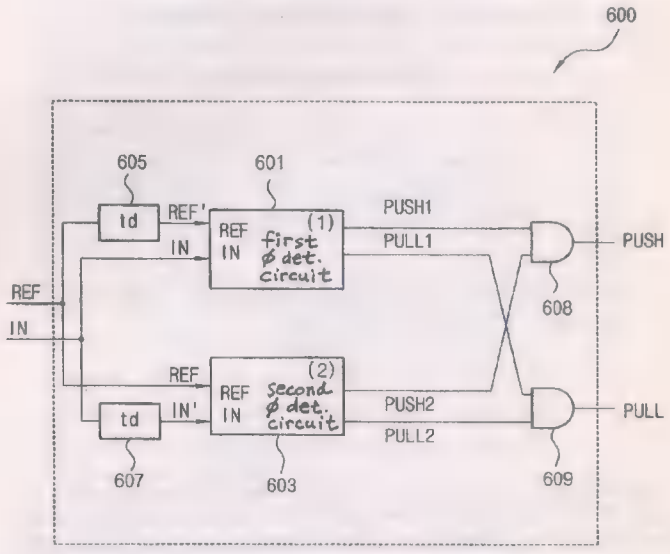
Current U.S. Class : **327/158**; 327/149; 327/161
Intern'l Class : H03L 007/06
Field of Search : 327/158, 161, 149, 152, 153, 2, 3, 5, 7, 12

References Cited

U.S. Patent Documents

5940344	Aug., 1999	Murai *et al.*	365/233.
6437618	Aug., 2002	Lee	327/158.
6614278	Sep., 2003	Kim *et al.*	327/263.

Abstract*—**A delay locked loop ***circuit with a novel structure for improving a jitter performance is disclosed. The delay locked loop ***circuit*** includes a delay ***circuit*** for receiving an input clock signal and generating ■ delayed output clock signal. The delay ***circuit*** has a predetermined minimum variable delay, and the output clock signal is delayed with respect to the input clock signal by a delay to be determined in accordance with a delay control signal inputted into the delay ***circuit***. Moreover, the delay locked loop ***circuit*** includes a phase determining block for receiving the input clock signal and the output clock signal, generating a phase pull signal when a phase of an input clock signal being delayed by a first predetermined time period leads ■ phase of the output clock signal, and generating ■ phase push signal when a phase of the input clock signal lags behind a phase of a delayed output clock signal delayed by a second predetermined time, and a delay control ***circuit*** for generating the delay control signal for controlling the delay ***circuit*** to reduce the delay when the phase pull signal is received from the phase determining block and to increase the delay when the phase push signal is received from the phase determining block. The delay control ***circuit*** does not change the delay of the delay ***circuit*** when neither the phase pull signal nor the phase push signal is received from the phase determining block.



6,762,644 July 13, 2004

Apparatus and Method for ■ Nested Transimpedance Amplifier

Inventor: Sutardja; Sehat (Cupertino, CA)
Assignee: Marvell International, Ltd. (Hamilton, BM)
Filed: February 6, 2002.

Current U.S. Class : **330/69**; 250/214A; 330/98; 330/99; 330/100; 360/77.02
Intern'l Class : H03F 003/45
Field of Search : 330/69, 98, 99, 100, 308, 260, 271 250/214 A 360/77.02

References Cited

U.S. Patent Documents

4535233	Aug., 1985	Abraham	250/214.
4564818	Jan., 1986	Jones	330/311.
4724315	Feb., 1988	Goerne	250/214.
4764732	Aug., 1988	Dion	330/59.
4772859	Sep., 1988	Sakai	330/308.
4914402	Apr., 1990	Dermitzakis *et al.*	330/308.
5010588	Apr., 1991	Gimlett	455/619.
5345073	Sep., 1994	Chang *et al.*	250/214.
5382920	Jan., 1995	Jung	330/308.
5532471	Jul., 1996	Khorramabadi *et al.*	250/214.
5646573	Jul., 1997	Bayruns *et al.*	330/59.
6037841	Mar., 2000	Tanji *et al.*	330/308.
6057738	May, 2000	Ku *et al.*	330/308.
6084478	Jul., 2000	Mayampurath	330/308.
6114913	Sep., 2000	Entrikin	330/308.
6122131	Sep., 2000	Jeppson	360/77.

Foreign Patent Documents

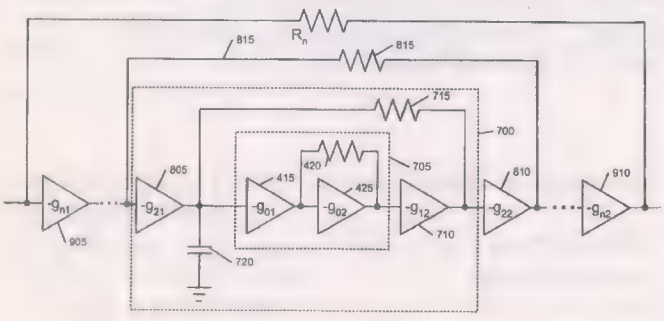
H6-61752	Mar., 1994	JP.	
406061752	Mar., 1994	JP	330/308.

Other References

You *et al.* "A Multistage Amplifier Topology with Nestal GM-C Compensation for Low-Voltage Application," IEEE International Solid-State Circuits Conference, 1997 Digest of Technical Papers, 44th ISSCC, Feb. 6-8, 1997, pp 348-39.

Holt "Electronic Circuits Digital and Analog" John Wiley & Sons, 1978, pp 423, 431, 436.

***Abstract*—A nested transimpedance amplifier (TIA) *circuit* includes a zero-order TIA having an input and an output. A first operational amplifier (opamp) has an input that communicates with the output of the zero-order TIA and an output. A first feedback resistor has one end that communicates with the input of the zero-order TIA and an opposite end that communicates with the output of the first opamp. A capacitor has one end that communicates with the input of the zero-order TIA. The gain-bandwidth product of the nested TIA is increased. Differential mode TIA's also have increased gain-bandwidth products.**



6,765,414 July 20, 2004

Low Frequency Testing, Leakage Control, and Burn-In Control for High-Performance Digital Circuits

Inventors: Keshavarzi; Ali (Portland, OR), Chatterjee; Bhaskar P. (Portland, OR), Krishnamurthy; Ram (Portland, OR), and Sachdev; Manoj (Ontario, CA).
Assignee: Intel Corporation (Santa Clara, CA)

Filed: September 17, 2002.

Current U.S. Class : **326/93**; 326/95; 326/112
Intern'l Class : H03K 019/096
Field of Search : 326/93, 95, 98, 16, 112, 119, 122

References Cited

U.S. Patent Documents

5619511	Apr., 1997	Sugisawa *et al.*	371/22.
5745499	Apr., 1998	Ong	371/21.
5748012	May, 1998	Beakes *et al.*	326/93.
5978944	Nov., 1999	Parvathala *et al.*	714/726.
6570407	May, 2003	Sugisawa *et al.*	326/93.

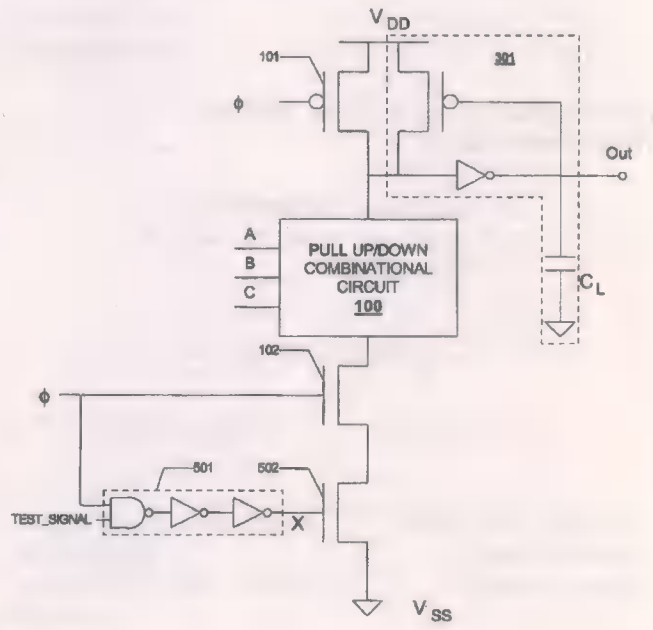
Other References

M. Shashaani *et al.*, A Design for Test Technique for High Performance Circuit Testing, IEEE International Test Conf., pp. 276-285, 1999.

V. D. Agrawal *et al.* High Performance Circuit Testing with Slow Speed Testers, Proc. of International Test Conf., pp. 302-310, 1995.

Y. Ye *et al.*, A New Technique for Standby Leakage Reduction in High Performance Circuits, Symp. on VLSI Circuits, p. 40, 1998.

***Abstract*—A technique is described to allow testing of high-speed *digital* circuits using lower speed testing equipment, to circuits to be placed into a sleep mode, and to allow burn-in testing of *digital* circuits with minimal overhead in terms of silicon area or performance.**



6,765,445 July 20, 2004

Digitally-Synthesized Loop Filter Circuit Particularly Useful for a Phase Locked Loop

Inventors: Perrott; Michael H. (Cambridge, MA), Baird; Rex T. (Nashua, TX), and Huang; Yunteng (Irvine, CA).
Assignee: Silicon Laboratories, Inc. (Austin, TX)
Filed: September 5, 2003.

Current U.S. Class : **331/17**; 331/11; 331/16; 331/18; 331/25; 375/247; 375/376
Intern'l Class : H03B 005/24
Field of Search : 331/17, 11, 16, 18, 25 375/376, 247

References Cited

U.S. Patent Documents

3968493	Jul., 1976	Last *et al.*	
4237423	Dec., 1980	Rhodes.	
4371974	Feb., 1983	Dugan.	
5005016	Apr., 1991	Schmidt *et al.*	
5027085	Jun., 1991	DeVito.	
5036294	Jul., 1991	McCaslin.	
5036298	Jul., 1991	Bulzachelli.	
5239561	Aug., 1993	Wong *et al.*	
5373255	Dec., 1994	Bray *et al.*	
5495512	Feb., 1996	Kovacs *et al.*	
5559841	Sep., 1996	Pandula.	
5631933	May, 1997	Chu *et al.*	375/376.
5734008	Mar., 1998	Shirasaki *et al.*	528/353.
5774023	Jun., 1998	Irwin	331/17.
5870003	Feb., 1999	Boerstler	331/17.
5892407	Apr., 1999	Ishii	331/17.
5942949	Aug., 1999	Wilson *et al.*	
5977838	Nov., 1999	Nagoya *et al.*	331/17.
5978426	Nov., 1999	Glover *et al.*	375/376.
5986512	Nov., 1999	Eriksson	331/16.
6008703	Dec., 1999	Perrott *et al.*	
6011815	Jan., 2000	Eriksson *et al.*	375/376.
6047029	Apr., 2000	Eriksson *et al.*	375/247.
6075388	Jun., 2000	Dalmia.	
6075416	Jun., 2000	Dalmia.	
6125158	Sep., 2000	Carson *et al.*	
6137372	Oct., 2000	Welland *et al.*	
6147567	Nov., 2000	Welland *et al.*	
6150891	Nov., 2000	Welland *et al.*	
6151152	Nov., 2000	Neary.	
6167245	Dec., 2000	Welland *et al.*	
6208211	Mar., 2001	Zipper *et al.*	
6580376	Jun., 2003	Perrott.	
6590426	Jul., 2003	Perrott.	
6643346	Nov., 2003	Pedrotti *et al.*	331/17.
6683506	Jan., 2004	Ye *et al.*	331/18.
6686805	Feb., 2004	Cyrusian	331/25.

Foreign Patent Documents

0590323	Apr., 1994	EP.
62-81813	Apr., 1987	JP.

Other References

Anderson, L. I. et al, "Silicon Bipolar Chipset for SONET/SDH 10Gb/s Fiber-Optic Communications Links," IEEE Journal of Solid-State Circuits, vol. 30, no. 3, Mar. 1995, pp. 210-218.

Belot, D. *et al.*, "A 3.3-V Power Adaptive 1244/622/155 Mbit/s Transciever for ATM, SONET/SDH," IEEE Journal of Solid-State Circuits, vol. 33, no. 7, Jul. 1998, pp. 1047-1058.

Gray, C. T. *et al.*, "A Sampling Technique and Its CMOS Implementation with 1 Gb/s Bandwidth and 25 ps Resolution," IEEE Journal of Solid-State Circuits, vol. 29, no. 3, Mar. 1994, pp. 340-349.

Guiterrez G. et al, "2.488 Gb/s Silicon Bipolar Clock and Data Recovery IC for SONET (OC-48)," IEEE 1998 Custom Integrated Circuits Conference, pp. 575-578.

Guiterrez, G. and Kong, S., "Unaided 2.5 Gb/s Silicon Bipolar Clock and Data Recovery IC," VIII-7, 1998 IEEE Radio Frequency Integrated Circuits Symposium, pp. 173-176.

Hogge, Charles R., Jr., "A Self Correcting Clock Recovery Circuit," IEEE Journal of Lightwave Technology, vol. LT-3, Dec. 1985, pp. 1312-1314, re-printed as pp. 249-251.

Hu, T. H. and Gray, P. R., "A Monolithic 480 Mb/s Parallel AGC/Decision/Clock-Recovery Circuit in 1.2-.mu.m CMOS," IEEE Journal of Solid-State Circuits, vol. 28, no. 12, Dec. 1993, pp. 1314-1320.

Jarman, David, "A Brief Introduction to Sigma Delta Conversion," Application Note AN9504, Intersil Corporation, May 1995, pp. 1-7.

Kawai, K. *et al.*, "A 557-mW, 2.5-Gbit/s SONET/SDH Regenerator-Section Terminating LSI Chip Using Low-Power Bipolar-LSI Design," IEEE Journal of Solid-State Circuits, vol. 34, no. 1, Jan. 1999, pp. 12-17.

Lee, T. H. and Bulzacchelli, J. F., "A 155-MHz Clock Recovery Delay- and Phase-Locked Loop," IEEE Journal of Solid-State Circuits, vol. SC-27, Dec. 1992, pp. 1736-1746, re-printed as pp. 421-430.

Lee, T. H. *et al.*, "A 2.5 V CMOS Delay-Locked Loop for an 18 Mbit, 500 Megabyte/s DRAM," IEEE Journal of Solid-State Circuits, vol. 29, no. 12, Dec. 1994, pp. 1491-1496.

Perrott, M. *et al.*, "A 27mW CMOS Fractional-N Synthesizer/Modulator IC," 1997 IEEE International Solid-State Circuits Conference, Session 22, Communications Building Blocks II, Paper SP 22.2, 1997 Digest of Technical Papers, vol. 40, pp. 366-367, 487.

Perrott, M. *et al.*, "A 27, W CMOS Fractional-N Synthesizer Using Digital Compensation for 2.5-Mb/s GFSK Modulation," IEEE Journal of Solid-States Circuits, vol. 32, no. 12, Dec. 1997, pp. 2048-2060.

Pottbacker, A. *et al.*, "A Si Bipolar Phase and Frequency Detector IC for Clock Extraction up to 8 Gb/s," IEEE Journal of Solid-State Circuits, vol. 27, no. 12, Dec. 1992, pp. 1747-1751.

Razavi, Behzad, "Design of Monolithic Phase-Locked Loops and Clock Recovery Circuits−−A Tutorial," Monolithic Phase-Locked Loops and Clock Recovery Circuits−−Theory and Design, ed. B. Razavi, IEEE Press, N.Y., 1996, pp. 1-39.

Walker, R. C. *et al.*, "A 10 Gb/s Si-Bipolar TX/RX Chipset for Computer Data Transmission," IEEE International Solid-State Circuits Conference, Session 19, Paper 19.1 Slide Supplement, 1998, pp. 19.1-1-19.1-11.

Walker, R. C. *et al.*, "A 1.5 Gb/s Link Interface Chipset for Computer Data Transmission," IEEE Journal on Selected Areas in Communications, vol. 9, no. 5, Jun. 1991, pp. 698-703.

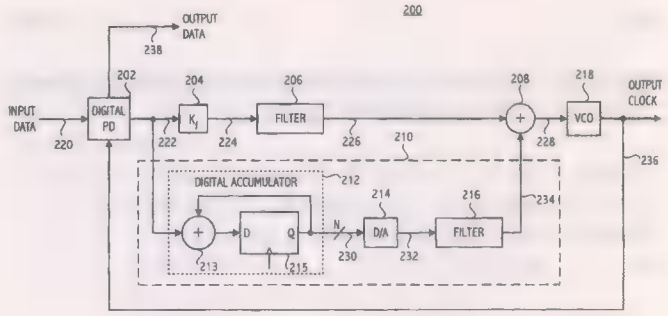
Weston, H. T. *et al.*, "A Submicrometer NMOS Multiplexer-Demultiplexer Chip Set for 622.08-Mb/s SONET Applications," IEEE Journal of Solid-State Circuits, vol. 27, no. 7, Jul. 1992, pp. 1041-1049.

Willingham, S. *et al.*, "An Integrated 2.5GHz .SIGMA..DELTA. Frequency Synthesizer with 5 .mu.s Settling and 2Mb/s Closed Loop Modulation," 2000 IEEE International Solid-State Circuits Conference, Session 12, Paper TP 12.3, pp. 200-201, 457.

Masaru Kokubo et al, "FA 15.2 : A Fast-Frequency-Switching PLL Synthesizer LSI with a Numerical Phase Comparator," IEEE International Solid-State Circuits Conference, New York, vol. 38, Feb. 1, 1995, pp. 260-261, 376.

Shayan, Y. R. *et al.*, "All Digital Phase-Locked Loop : Concepts, Design and Applications," IEEE Proceedings-F/Radar and Signal Processing 136, Stevenage, Herts, GB, vol. 136, no. 1, Part F, Feb. 1, 1989, pp. 53-56.

Abstract*—**In a feedback system such as a PLL, the integrating function associated with a loop filter capacitor is instead implemented digitally and is easily implemented on the same integrated ***circuit die as the PLL. There is no need for either an external loop filter capacitor nor for a large loop filter capacitor to be integrated on the same integrated ***circuit*** die as the PLL. In a preferred embodiment, an analog phase detector is utilized whose phase error output signal is delta-sigma modulated to encode the magnitude of the phase error using a digital (i.e., discrete-time and discrete-value) signal. This digital phase error signal is "integrated" by a digital integration block including, for example, a digital accumulator, whose output is then converted to an analog signal, optionally combined with a loop feed-forward signal, and then conveyed as a control voltage to the voltage-controlled oscillator. The equivalent "size" of the integrating capacitor function provided by the digital integration block may be varied by increasing or decreasing the bit resolution of circuits within the digital block. Consequently, an increasingly larger equivalent capacitor may be implemented by adding additional digital stages, each of which requires a small incremental integrated ***circuit*** area. The power dissipation of the digital integration block is reduced by incorporating a decimation stage to reduce the required operating frequency of the remainder of the digital integration block.



6,766,153 July 20, 2004

Dynamic Automatic Gain Control Circuit Employing Kalman Filtering

Inventors: Kozak; Marian (Holon, IL) and Raphaeli; Dan (Kfar Saba, IL).
Assignee: Itran Communications Ltd. (Beersheva, IL)
Filed: April 2, 2001.

Current U.S. Class : **455/232.1**; 327/205; 330/75; 455/250.1
Intern'l Class : H04B 007/00
Field of Search : 455/232.1, 234.1, 236.1, 240.1, 245.1, 250.1, 255, 136, 138, 260 330/75, 107, 103, 254, 278 327/154, 155, 205, 323

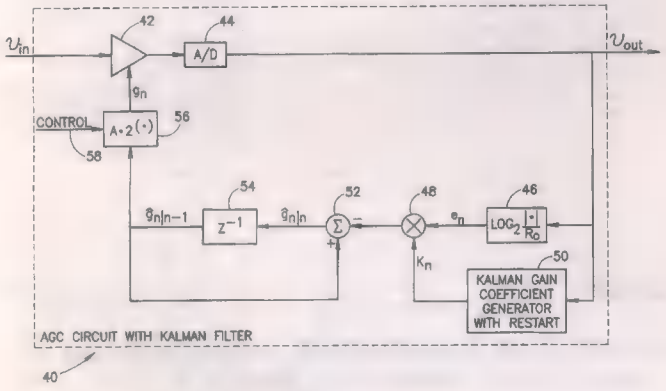
References Cited

U.S. Patent Documents

3943286	Mar., 1976	Tsurushima	179/1.
4123718	Oct., 1978	Lampert *et al.*	325/474.
4300102	Nov., 1981	Inoue	330/254.
4349883	Sep., 1982	Doljack	364/551.
4353036	Oct., 1982	Hoover	330/264.
4499429	Feb., 1985	Sugimoto	330/254.
4677392	Jun., 1987	Yang	330/284.
4797632	Jan., 1989	Guery	330/285.
4839573	Jun., 1989	Wise	318/615.
4878031	Oct., 1989	Main	330/254.
4972163	Nov., 1990	Van Der Plas	331/12.
5051707	Sep., 1991	Fujita	330/279.
5117201	May, 1992	Luther	330/279.
5194822	Mar., 1993	Bureau *et al.*	330/129.
5307026	Apr., 1994	Mucke	330/283.
5459383	Oct., 1995	Sidman *et al.*	318/611.
5592176	Jan., 1997	Vickers *et al.*	
5668501	Sep., 1997	Venes	330/254.
5684431	Nov., 1997	Gilbert *et al.*	330/254.
5734974	Mar., 1998	Callaway, Jr. *et al.*	455/234.
5742206	Apr., 1998	Ishida	330/284.
5774829	Jun., 1998	Cisneros *et al.*	

5838395	Nov., 1998	Brilka	348/726.
5867222	Feb., 1999	Norris *et al.*	348/528.
5952880	Sep., 1999	Voorman *et al.*	330/254.
6018554	Jan., 2000	Glover	375/345.
6044253	Mar., 2000	Tsumura	455/234.
6107878	Aug., 2000	Black	330/129.
6111710	Aug., 2000	Feyh *et al.*	360/46.
6144233	Nov., 2000	Maruyama *et al.*	327/77.
6166850	Dec., 2000	Roberts *et al.*	359/341.
6169452	Jan., 2001	Popescu *et al.*	330/254.
6169638	Jan., 2001	Morling	360/46.
6181201	Jan., 2001	Black	330/129.
6195028	Feb., 2001	Fredrickson *et al.*	341/132.
6285863	Sep., 2001	Zhang.	
6563891	May, 2003	Eriksson *et al.*	375/345.

Abstract*—A *novel and useful apparatus for and method of automatic gain control (AGC) using Kalman filtering and hysteresis. A nonlinear, time-variant loop filter such as a Kalman filter is employed in the feedback loop of an AGC ***circuit***. The ***circuit*** is able to transition quickly and make fast adaptations to new levels of the input signal by use of a restart mechanism used to dynamically modify the gain of the loop filter thus enabling the AGC ***circuit*** to quickly adapt to changes in the signal level of the input. An AGC ***circuit*** incorporating a hysteresis ***circuit*** in the feedback loop is also disclosed.



6,778,117 August 17, 2004

Local Oscillator and Mixer for a Radio Frequency Receiver and Related Method

Inventor: Johnson; Richard A. (Buda, TX)
Assignee: Silicon Laboratories, Inc. (Austin, TX)
Filed: February 28, 2003.

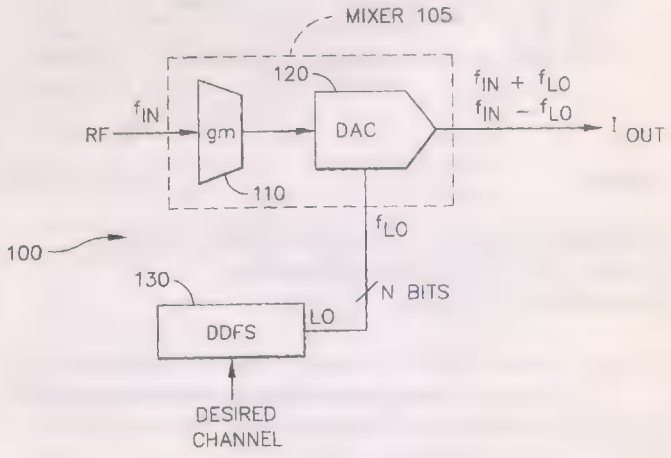
Current U.S. Class :	**341/144**; 341/118
Intern'l Class :	H03M 001/66
Field of Search :	341/144, 145, 136, 118 330/253, 269 375/376, 362

References Cited

U.S. Patent Documents

4866261	Sep., 1989	Pace	341/138.
5442352	Aug., 1995	Jackson.	
5495512	Feb., 1996	Kovacs *et al.*	375/376.
5737035	Apr., 1988	Rotzoll.	
6172569	Jan., 2001	McCall *et al.*	330/303.
6177964	Jan., 2001	Birleson *et al.*	
6377315	Apr., 2002	Carr *et al.*	
6600373	Jul., 2003	Bailey *et al.*	330/260.
2003/0223525	Dec., 2003	Momtaz *et al.*	375/376.

Abstract*—A *circuit (100) is adapted for use in a radio frequency receiver and includes a transconductance amplifier (110), a direct digital frequency synthesizer (130), and a digital-to-analog converter (DAC) (120). The transconductance amplifier (110) has an input terminal for receiving a radio frequency signal, and an output terminal for providing a current signal. The direct digital frequency synthesizer (130) has an output terminal for providing a digital local oscillator signal at a selected frequency. The DAC (120) has a first input terminal coupled to the output terminal of the transconductance amplifier (110), a second input terminal coupled to the output terminal of the direct digital frequency synthesizer (130), and an output terminal for providing an output signal.



6,778,126 August 17, 2004

Structures and Methods That Improve the Linearity of Analog-to-Digital Converters With Introduced Nonlinearities

Inventor: Ali; Ahmed Mohamed Abdelatty (Greensboro, NC)
Assignee: Analog Devices, Inc. (Norwood, MA)
Filed: November 21, 2002.

Current U.S. Class :	**341/156**; 341/118; 341/140
Intern'l Class :	H03M 001/06; H03M 001/12
Field of Search :	341/156, 118, 140, 150, 155, 138, 139, 120, 161 708/7

References Cited

U.S. Patent Documents

4654815	Mar., 1987	Marin *et al.*	708/7.
5635937	Jun., 1997	Lim *et al.*	
6184809	Feb., 2001	Yu.	
6259392	Jul., 2001	Choi *et al.*	341/150.
6369744	Apr., 2002	Chuang.	
6373424	Apr., 2002	Soenen.	
6445319	Sep., 2002	Bugeja	
6545628	Apr., 2003	Aram	341/155.
5963156	Oct., 1999	Lewicki *et al.*	327/94.
6337651	Jan., 2002	Chiang *et al.*	341/161.
6362688	Mar., 2002	Au *et al.*	330/253.

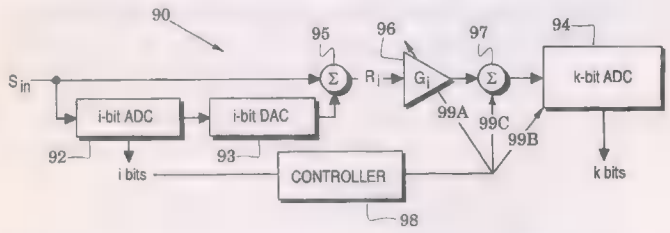
Foreign Patent Documents

0 380 152 Aug., 1990 EP.

Other References

International Search Report issued Jul. 24, 2003 for Appln. no. PCT/US01/41576, 4 pages.

***Abstract*—**Analog-to-digital converter (ADC) structures and methods are provided that reduce an initial converter nonlinearity by introducing an inverse nonlinearity into the converter's response that is substantially the inverse of the initial converter nonlinearity. In a pipelined ADC embodiment, for example, upstream converter stages are selected that generate an upstream digital code which defines sufficient upstream code words to designate respective segments of the inverse nonlinearity. In response to each of the upstream code words, the conversion gain of the remaining downstream converter stages is then sufficiently adjusted to insert the inverse nonlinearity into the converter response.



6,781,451 August 24, 2004

Switched-Capacitor, Common-Mode Feedback Circuit for a Differential Amplifier Without Tail Current

Inventors: Kwan; Tom W. (Cupertino, CA), Duncan; Ralph (Laguna Beach, CA), and Singor; Frank W. (Laguna Beach, CA).
Assignee: Broadcom Corporation (Irvine, CA)
Filed: April 30, 2003.

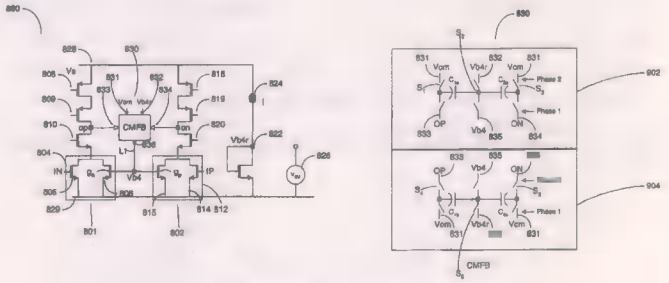
Current U.S. Class : **330/9**; 330/258
Intern'l Class : H03F 001/02
Field of Search : 327/9, 258, 254, 278, 290, 291, 337, 554, 74, 75, 307 330/253, 258, 9, 254, 278, 290, 291

References Cited

U.S. Patent Documents

4849662	Jul., 1989	Hoelberg *et al.*	327/554.
4992755	Feb., 1991	Seevinck *et al.*	330/253.
5710563	Jan., 1998	Vu *et al.*	341/161.
5864257	Jan., 1999	Rothenberg	330/253.
5894284	Apr., 1999	Garrity *et al.*	327/337.

Abstract*—**Provided is a switched capacitor feedback ***circuit including two or more input ports configured to receive a corresponding a number of input signals and at least one output port. The output port is configured to output an adjusting signal. The input signals includes a number of primary signals and two or more reference signals that are associated with a first timing phase of operation. The adjusting signal is produced based upon a comparison between the primary signals the reference signals. Also provided is a pair of active devices having gates coupled together and structured to receive the adjusting signal. The active devices are configured to provide a gain to the adjusting signal in accordance with a predetermined gain factor, and facilitate an adjustment to the number of primary signals based upon the gain during a second timing phase of operation.



6,791,413 September 14, 2004

Variable Gain Amplifier With a Gain Exhibiting Linear in dB Characteristic Relative to a Control Voltage

Inventors: Komurasaki; Hiroshi (Hyogo, JP), Satoh; Hisayasu (Hyogo, JP), Hosoda; Kinya (Saitama, JP), Hyogo; Akira (Chiba, JP), and Sekine; Keitaro (Tokyo, JP).
Assignee: Renesas Technology Corp. (Tokyo, JP)
Filed: March 10, 2003.

Foreign Application Priority Data

Sep 10, 2002[JP] 2002-264124

Current U.S. Class : **330/254**; 330/253
Intern'l Class : H03F 003/45
Field of Search : 330/254, 253, 257 327/359

References Cited

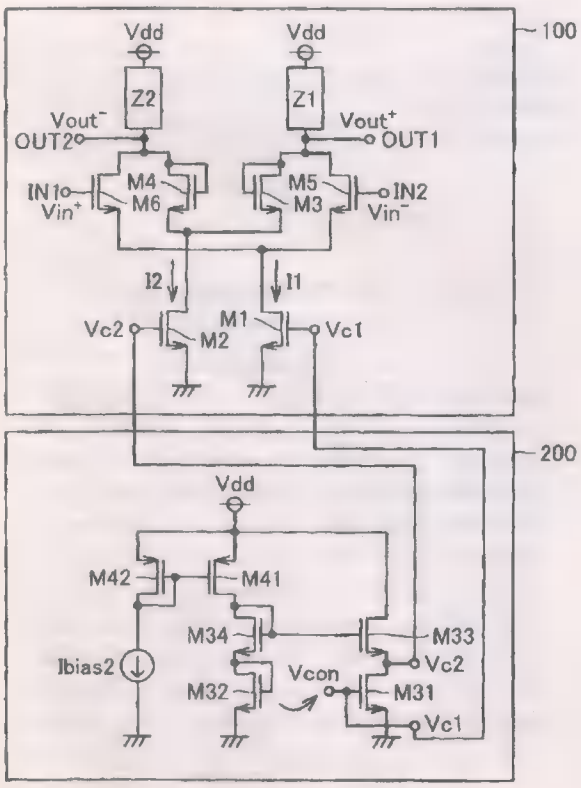
U.S. Patent Documents

6163215	Dec., 2000	Shibata *et al.*	330/254.
6552611	Apr., 2003	Yamamoto	330/253.
6566951	May, 2003	Merrigan *et al.*	330/254.

Other References

Po-Chiun Huang *et al.*, "A 3.3-V CMOS Wideband Exponential Control Variable-Gain-Amplifier," Circuits and Systems, 1998, ISCAS '98, Proceedings of the 1998 IEEE International Symposium on May 31-Jun. 3, 1998, vol. 1, pp. 285-288.

Abstract*—A variable gain amplifier is configured of an amplification *circuit and a control ***circuit*** controlling a gain of the amplification ***circuit***. The amplification ***circuit*** has first and second MOS transistors identical in characteristics and having respective sources connected to a first fixed potential. The amplification ***circuit*** has a differential gain proportional to a square root of a ratio between a current flowing through the first MOS transistor and a current flowing through the second MOS transistor. The control ***circuit*** applies a potential corresponding to a constant voltage plus a control voltage to a gate of the first MOS transistor and a potential corresponding to the constant voltage minus the control voltage to a gate of the second MOS transistor.



6,791,431 September 14, 2004

Compact Balun With Rejection Filter for 802.11a and 802.11b Simultaneous Operation

Inventor: De Flaviis; Franco (Irvine, CA)
Assignee: Broadcom Corporation (Irvine, CA)

Filed: September 3, 2002.

Current U.S. Class : **333/26**; 333/116
Intern'l Class : H01P 005/10
Field of Search : 333/25, 26, 116, 109, 112, 115

References Cited

U.S. Patent Documents

4375699	Mar., 1983	Hallford	455/327.
5455545	Oct., 1995	Garcia	333/26.
6018277	Jan., 2000	Vaisanen	333/26.
6300919	Oct., 2001	Mehen *et al.*	343/895.
6515556	Feb., 2003	Kato *et al.*	333/116.

Foreign Patent Documents

2144985 Mar., 1973 EP

Other References

Piernes B. *et al.*, "Improvement of the Design of 180 DEG Rat-Race Hybrid", Electronics Letters, GB, vol. 36, No. 12, pp. 1035-1036 (Jun. 2000).

Settaluri Raghu K. et al., "Compact Folded Line Rat-Race Hybrid Couplers", IEEE Microwave Guided Wave Lett; IEEE Translations on Microwave and Guided Wave Letters, Feb. 2000, IEEE, vol. 10, No. 2, pp. 61-63.

Settaluri R.K. et al., "Design of Compact Multi-Level Folded-Line RF Couplers", 1999 IEEE MTT-S International Microwave Symposium Digest, IEEE Transactions on Microwave Theory and Techniques, vol. 47, No. 12, pp. 2331-2339 (Dec. 1999).

Matsuura, H. et al., "Monolithic Rat-Race Mixers for Millimeter Waves", IEEE, pp. 101-104 (Jul. 2004).

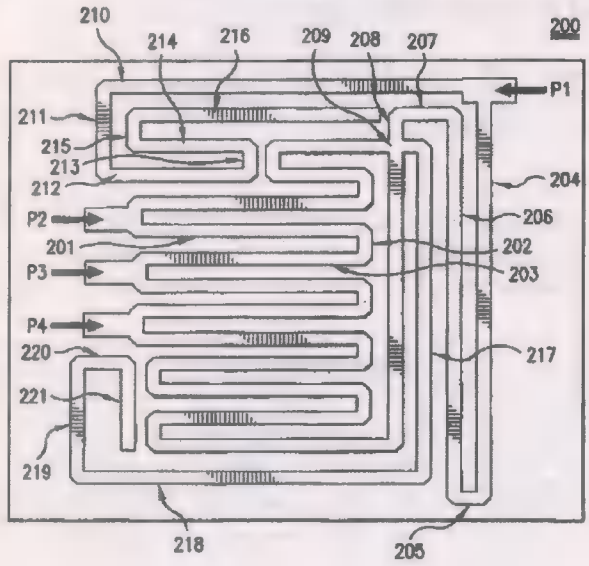
Johnson, K.M., "X-Band Integrated Circuit Mixer with Reactively Terminated Image", Transactions n Microwave Theory and Techniques, IEEE, vol. 16, No. 7, pp. 388-397 (Jul. 1968).

Copy of European Search Report issued Jan. 13, 2004 for Appl. No. EP/03019915, 5 pages.

Copy of European Search Report issued Jan. 13, 2004 for Appl. No. EP/03019914.4, 4 pages.

***Abstract*—A balancing/unbalancing (balun) structure for operating at frequency f_1 includes a microstrip printed *circuit* board (PCB). A balun on the PCB includes two input ports are coupled to a differential signal. An isolated port is connected to ground through a matched resistance. An output port is coupled to a single-ended signal corresponding to the differential signal. A plurality of traces on the PCB connect the two input ports, the load connection port and a tap point to the output port. A f_2 rejection filter on the PCB is wrapped around the balun and includes a first folded element with a transmission length of $\lambda_2/4$**

and connected to the output port. A second folded element has a transmission length of $\lambda_2/4$ and connected to the tap point. A third folded element connects the tap point to the output port and has a transmission length of $\lambda_2/4$.



6,794,914 September 21, 2004

Non-Volatile Multi-Threshold CMOS Latch With Leakage Control

Inventors: Sani; Mehdi Hamidi (San Diego, CA) and Uvieghara; Gregory A. (San Diego, CA).
Assignee: Qualcomm Incorporated (San Diego, CA)
Filed: May 24, 2002.

Current U.S. Class : **327/202**; 327/203
Intern'l Class : H03K 003/289; H03K 003/356
Field of Search : 327/199, 200-203, 208-212, 218 326/40, 46

References Cited

U.S. Patent Documents

5982211	Nov., 1999	Ko	327/202.
6246265	Jun., 2001	Ogawa	326/95.
6304123	Oct., 2001	Bosshart	327/212.
6310491	Oct., 2001	Ogawa	326/46.
6538471	Mar., 2003	Stan et al.	326/46.
6566927	May, 2003	Park et al.	327/211.
2002/0080663	Jun., 2002	Kameyama et al.	

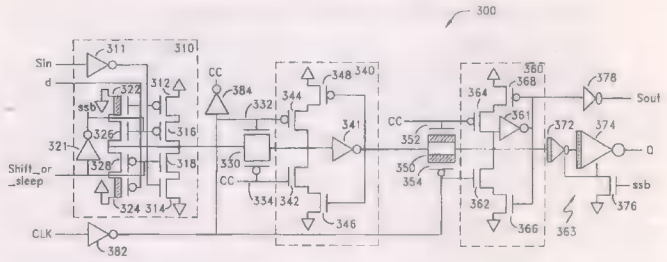
Other References

1-V Power Supply High-Speed Digital Circuit Technology with Multithreshold-Voltage CMOS : IEEE Journal of Solid-State Circuits. vol. 30 : 8 (1995).

SSCC96. Session 10/Low Power & Communication Signal Processing. Paper FA 10.4.

A 1-V High-Speed MTCMOS Circuit Scheme for Power-Down Application Circuits : IEEE Journal of Solid-State Circuits. vol. 32 : 6 (1997).

***Abstract*—**An integrated circuit including a Multi-Threshold CMOS (MTCMOS) latch combining low voltage threshold CMOS circuits with high voltage threshold CMOS circuits. The low voltage threshold circuits including a majority of the circuits in the signal path of the latch to ensure high performance of the latch. The latch further including high voltage threshold circuits to eliminate leakage paths from the low voltage threshold circuits when the latch is in a sleep mode. A single-phase latch and a two-phase latch are provided. Each of the latches is implemented with master and slave registers. Data is held in either the master register or the slave register depending on the phase or phases of the clock signals. A multiplexer may alternatively be implemented prior to the master latch for controlling an input signal path during sleep and active modes of the latch and for providing a second input signal path for test.



6,795,007 September 21, 2004

Circuits and Methods for a Variable Over Sample Ratio Delta-Sigma Analog-to-Digital Converter

Inventor: Mayes; Michael Keith (Campbell, CA)
Assignee: Linear Technology Corporation (Milpitas, CA)
Filed: October 28, 2003.

Current U.S. Class : **341/143**; 341/61; 341/155
Intern'l Class : H03M 003/00
Field of Search : 341/61, 77, 120, 131, 143, 155

References Cited

U.S. Patent Documents

4943807	Jul., 1990	Early et al.	341/120.
4972436	Nov., 1990	Halim et al.	375/247.
5144308	Sep., 1992	Norsworthy	341/131.
5187482	Feb., 1993	Tiemann et al.	341/143.
5757299	May, 1998	Noro et al.	341/143.
6124815	Sep., 2000	Lee et al.	341/143.
6140950	Oct., 2000	Oprescu	341/143.
6169506	Jan., 2001	Oprescu et al.	341/143.
6208279	Mar., 2001	Oprescu	341/143.
6285306	Sep., 2001	Zrilic	341/143.
6337645	Jan., 2002	Pflaumer	341/143.

Other References

M. P. Doandio, "CIC Filter Introduction," downloaded from the Internet at http://www.dspguru.com/info/tutor/cic.htm, Jul. 18, 2000.

"Cascaded Integrator-Comb (CIC) Filter v. 1.0," Product Specification, LogiCore. Xilinx, Inc., Mar. 02, 2001.

G. Noriega, "Sigma-Delta A/D Converters-Audio and Medium Bandwidths," Technical Notes-DT3, RMS Instruments, downloaded from the Internet at http://www.rmsinst.com/dt3.htm, Feb. 1996.

"An Overview of Data Converters," Application Note AN100, Philips Semiconductors, Dec. 1991.

S. Park, "Principles of Sigma-Delta Modulation for Analog-to-Digital Converters," Communications Applications Manual. APR8, Motorola, DL411D/REV 1, 1993.

E. Dijkstra, et al., "On the Use of Modulo Arithmetic Comb Filters in Sigma Delta Modulators," IEEE Proc. ICASSP, pp. 2001-2004, Apr. 1988.

B. E. Boser *et al.*, "The Design of Sigma-Delta Modulation Analog-to-Digital Converters," IEEE Journal of Solid State Circuits, vol. 23, pp. 1298-1303, Dec. 1988.

J. C. Candy, "Decimation for Sigma Delta Modulation," IEEE Transactions on Communications, vol. COM-34, pp. 72-76, Jan. 1986.

J. C. Candy, et al., "Oversampling Delta-Sigma Data Converters--Theory, Design, and Simulation," IEEE Press, NY, pp. 1-25, 1992.

"ADC0820 : 8-Bit High Speed .mu.P Compatible A/D Converter With Track/Hold Function," datasheet, NAtional Semiconductor, Jun. 1999.

"LTC1410 : 12-Bit, 1.25 Msps Sampling A/D Converter with Shutdown," datasheet, Linear Technology, 1995.

ADC0801//ADC0802/ADC0803/ADC0804/ADC0805 : 8-Bit .mu.P Compatible A/D Converters, datasheet, National Semiconductor, Nov. 1999.

"AD650 : Voltage-to-Frequency and Frequency-to-Voltage Converter," datasheet, Linear Technology, 1995.

"LM231A/LM231/LM331A/LM331 : Precision Voltage-to-Frequency Converters," datasheet, National Semiconductor, Jun. 1999.

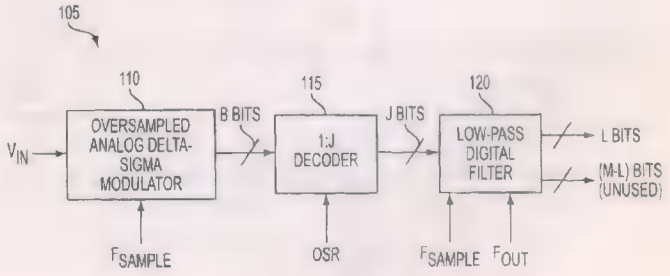
"ALD500AU/ALD500A/ALD500 : PRecision Integrating Analog Processor," datasheet, Advanced Linear Devices, Inc., 1999.

"AD1170 : High Resolution, Programmable Integrating A/D Converter," datasheet, Rev. A, Analog Devices, Aug. 1999.

"LTC2400 : 24-Bit .mu.Power No Latency .DELTA..SIGMA..TM. ADC in SO-8," datasheet, Linear Technology, 1998

"LTC2410 : 24-Bit No Latency .DELTA..SIGMA..TM. ADC with Differential Input and Differential Reference," datasheet, Linear Technology, 2000.

Abstract*—**Circuits and methods for a delta-sigma ***analog-to-***digital*** converter having a variable oversample ratio to produce a constant fullscale output at reduced circuit complexity, die area, and power dissipation are provided. The circuits and methods consist of scaling the ***digital*** input to the ***digital*** filter with a decoder whose size depends on the number of oversample ratios allowed by the ***analog***-to-***digital*** converter. The ***digital*** filter is implemented as a comb filter having a cascade of N integrators and N differentiators, where N is the order of the digital filter. The size of the differentiators is equal to the number of bits used as output for the ***analog***-to-***digital*** converter, which is smaller than the size of the integrators and the number of bits produced by the ***digital*** filter.



6,801,099 October 5, 2004

Methods for Bi-Directional Signaling

Inventor: Stark; Donald C. (Los Altos Hills, CA)
Assignee: Rambus Inc. (Los Altos, CA)
Filed: July 16, 2003.

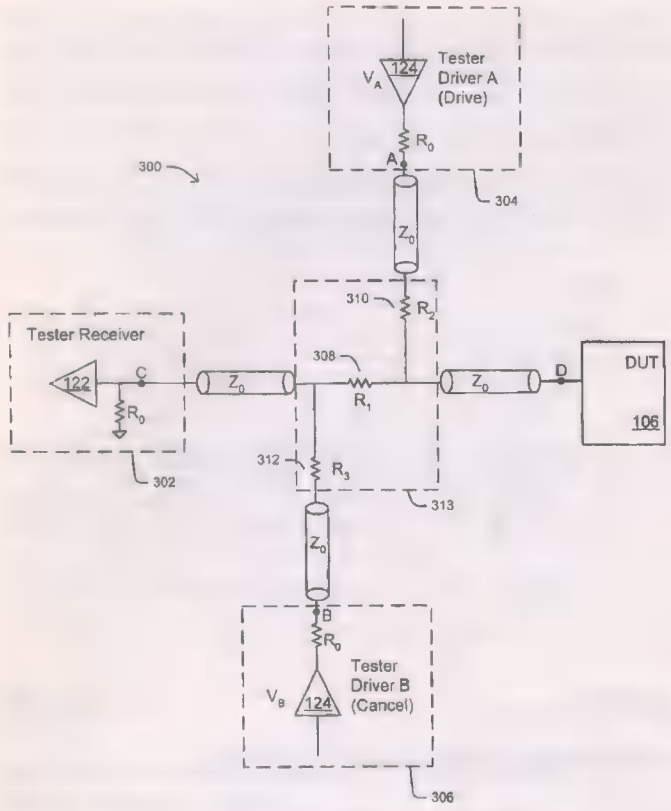
Current U.S. Class :	**333/130**; 324/329; 324/759; 324/765; 370/282
Intern'l Class :	G01R 031/26; H03H 007/38
Field of Search :	333/130, 324/765, 329, 759 370/282

References Cited

U.S. Patent Documents

5719856	Feb., 1998	May	370/282.
6452428	Sep., 2002	Mooney et al.	327/108.

Abstract*—**Improved methods and apparatuses are provided for conducting bi-directional signaling and testing. The outputs of at least two driver circuits are connected to a resistive network. The output signals from the driver circuits are combined through the resistive network to produce a resultant signal that is an attenuated version of at least one of the output signals. The resistive network and the driver circuits are configured such that the resultant signal is provided to an output node of the resistive network but not to an input node of the resistive network. An input/output node of an external ***circuit is connected to the input node of the resistive network, wherein the external ***circuit*** is configured to receive the resultant signal and output an external signal. An input node of a receiver ***circuit*** is connected to the output node of the resistive network. The resultant signal is then simultaneously provided to the external ***circuit*** and the external signal to the receiver ***circuit***, bi-directionally through the resistive network.



6,806,744 October 19, 2004

High Speed Low Voltage Differential to Rail-to-Rail Single Ended Converter

Inventors: Bell; Marshall J. (Chandler, AZ), Cooper; David B. (Chandler, AZ), and Kozisek; James (Fort Collins, CO).
Assignee: National Semiconductor Corporation (Santa Clara, CA)
Filed: October 3, 2003.

Current U.S. Class : **327/70**; 327/53; 327/65
Intern'l Class : H03K 005/22
Field of Search : 327/52, 53, 65, 66, 70

References Cited

U.S. Patent Documents

4539489	Sep., 1985	Vaughn	327/206.
6429735	Aug., 2002	Kuo et al.	327/563.
6433602	Aug., 2002	Lall et al.	327/205.
6512400	Jan., 2003	Forbes	327/66.

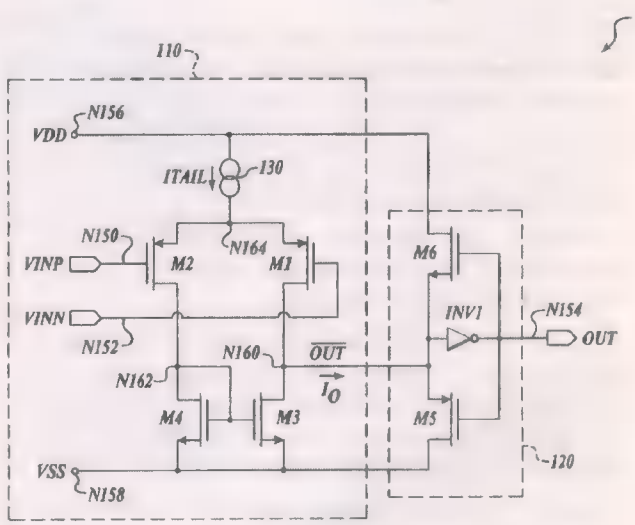
Other References

H. Traff, "Novel Approach to High Speed CMOS Current Comparators," Electronics Letters, vol. 28, No. 3, 30th Jan. 1992, (3 pgs.).

William Redman-White, "A High Bandwidth Constant g.sub.m and Slew-Rate Rail-to-Rail CMOS Input Circuit and its Application to Analog Cells for Low Voltage VLSI Systems,"IEEE vol. 32, No. 5, May 1997 (12 pgs.).

L. Ravezzi et al., "Simple High-Speed CMOS Current Comparator," Electronics Letters, vol. 33, No. 22, 23rd Oct., 1997 (2 pages).

Abstract*—**A method and system is arranged to convert a differential low-voltage input signal (e.g. LVDS or RSDS) into a single-ended output signal. An operational transconductance amplifier (OTA) is configured to convert the input signal into a current. A transimpedance stage is configured to convert the current into the single-ended output signal. The voltage associated with the output of the OTA corresponds to approximately VDD/2. The transimpedance stage comprises an inverter ***circuit, a p-type transistor, and an n-type transistor. The transistors are arranged in a negative feedback configuration with the inverter. The single-ended output signal has a voltage swing that approximately corresponds to the sum of the V.sub.GS of the n-type transistor and the V.sub.GS of the p-type transistor. The output signal may be buffered by additional circuits such as an inverter, a Schmitt, as well as others.



6,806,787 October 19, 2004

Varactor Folding Technique for Phase Noise Reduction in Electronic Oscillators

Inventors: Gomez; Ramon Alejandro (San Juan, CA), Burns; Lawrence M. (Luguna Mills, CA), and Kral; Alexandre (Laguna Niguel, CA).
Assignee: Broadcom Corporation (Irvine, CA)
Filed: March 25, 2003.

Current U.S. Class : **331/179**; 331/117FE; 331/175; 331/177V
Intern'l Class : H03B 005/08; H03B 005/12
Field of Search : 331/36 C, 117 R, 117 FE, 117 D, 175, 177 R, 177 V, 179

References Cited

Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proceedings of the IEEE--Frequency Stability, manuscript received : Dec. 10, 1965, revised : Dec. 29, 1965, Feb., 1996, pp. 329-330.

Ravazi, "Oscillators," RF Microelectronics, Chapter 7, pp. 206-246, Prentice Hall PTR 1998.

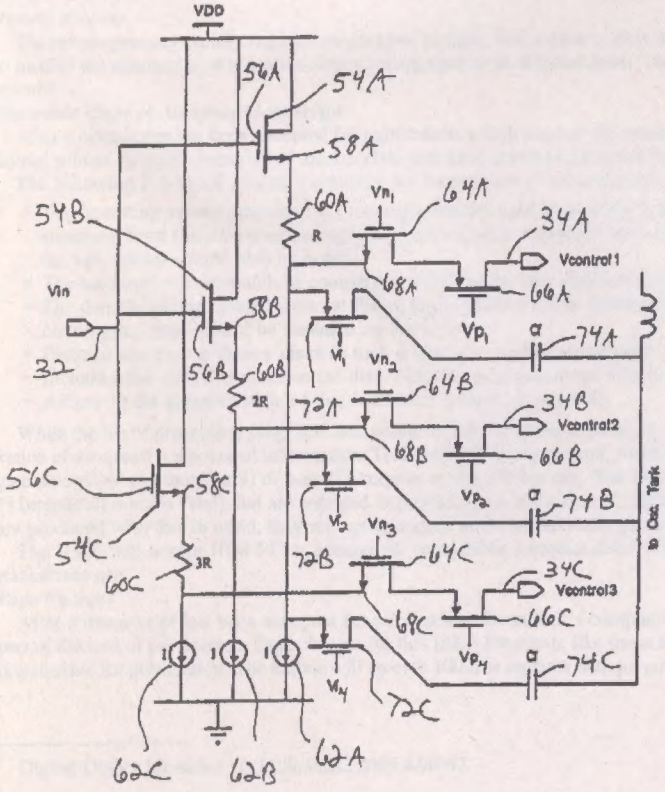
Rohde, "Microwave and Wireless Synthesizers--Theory and Design," pp. 567-572, John Wiley & Sons, Inc. 1997.

Copy of International Search Report For International Application no. PCT/US00/34095, filed Dec. 14, 2000.

Hajimiri *et al.*, "A General Theory of Phase Noise in Electrical Oscillators," IEEE Journal of Solid-State Circuits, vol. 33, no. 2, Feb., 1998, 179-194.

Kral *et al.*, "RF-CMOS Oscillators with Switched Tuning," IEEE 1998 Custom Integrated Circuits Conference, May 11-14, 1998, pp. 26.1.1-26.1.4, pp. 555-558.

***Abstract*—A varactor folding technique reduces noise in controllable electronic oscillators through the use of a series of varactors having relatively small capacitance. A folding *circuit* provides control signals to the varactors in a sequential manner to provide a relatively smooth change in the total capacitance of the oscillator. Consequently, effective control of the oscillator is achieved with accompanying reductions in oscillator noise such as flicker noise.**



6,809,425 October 26, 2004

Integrated Circuit With a Reprogrammable Nonvolatile Switch Having a Dynamic Threshold Voltage (VTH) for Selectively Connecting a Source for a Signal to a Circuit

Inventors: Chen; Bomy (Cupertino, CA), Nojima; Isao (Los Altos, CA), and Nguyen; Hung Q. (Fremont, CA).
Assignee: Silicon Storage Technology, Inc. (Sunnyvale, CA)

Filed: August 15, 2003.

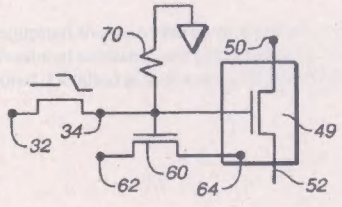
Current U.S. Class : **257/315**; 257/901
Intern'l Class : H01L 027/088
Field of Search : 257/901, 315

References Cited

U.S. Patent Documents

5029130	Jul., 1991	Yeh	365/185.
6232893	May, 2001	Cliff et al.	341/78.

***Abstract*—A nonvolatile reprogrammable switch for use in a PLD or FPGA has a nonvolatile memory cell connected to the gate of an MOS transistor, which is in a well, with the terminals of the MOS transistor connected to the source of the signal and to the *circuit*. The nonvolatile memory cell is of a split gate type having a first region and a second region, with a channel therebetween. The cell has a floating gate positioned over a first portion of the channel, which is adjacent to the first region and a control gate positioned over a second portion of the channel, which is adjacent to the second region. The second region is connected to the gate of the MOS transistor. The cell is programmed by injecting electrons from the channel onto the floating gate by hot electron injection mechanism. The cell is erased by Fowler-Nordheim tunneling of the electrons from the floating gate to the control gate. As a result, no high voltage is ever applied to the second region during program or erase. In addition, a MOS FET transistor has a terminal connected to the well, and another end to a voltage source, with the gate connected to the nonvolatile memory cell. The switch also has a *circuit* element connecting the gate of the MOS transistor to a voltage source. The threshold voltage of the well can be dynamically changed by turning on/off the MOS FET transistor.**



6,809,676 October 26, 2004

Method and System for VCO-Based Analog-to-Digital Conversion (ADC)

Inventors: Younis; Ahmed (Austin, TX), Hassoun; Marwan M. (Austin, TX), and Robinson; Moises E. (Austin, TX).
Assignee: Xilinx, Inc. (San Jose, CA)
Filed: August 20, 2002.

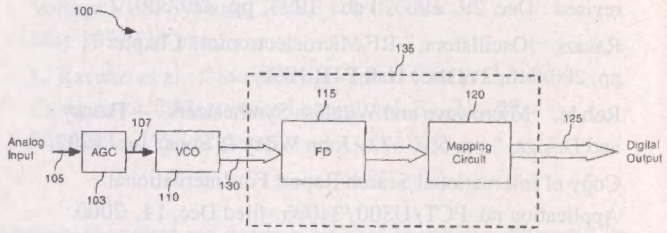
Current U.S. Class : **341/157**; 375/375
Intern'l Class : H03M 001/60; H03D 003/24
Field of Search : 341/139, 142, 157, 161, 152, 166 375/148, 316, 375, 376

References Cited

U.S. Patent Documents

5005016	Apr., 1991	Schmidt *et al.*	341/142.
5189420	Feb., 1993	Eddy *et al.*	341/157.
RE34899	Apr., 1995	Gessaman *et al.*	341/157.
5511100	Apr., 1996	Lundberg *et al.*	375/376.
5671252	Sep., 1997	Kovacs *et al.*	375/316.
5796358	Aug., 1998	Shih *et al.*	341/139.
6192094	Feb., 2001	Herrmann *et al.*	375/375.
6278725	Aug., 2001	Rouphael *et al.*	375/148.
6677879	Jan., 2004	Nix *et al.*	341/161.

***Abstract*—A VCO (110) can be configured to convert an *analog* input signal (105) to a *digital* output signal (125). In accordance with the inventive arrangements, the VCO can convert the *analog* input signal to at least one intermediate signal (130) having a frequency dependent on the *analog* input signal. A frequency detector (115) can be configured to determine a frequency of at least one intermediate signal. Subsequently, a mapping circuit (120) can be configured to map the determined frequency of the at least one intermediate signal to an output value representing the *digital* output signal (125).**



INFORMATION FOR AUTHORS

Content

The IEEE JOURNAL OF SOLID-STATE CIRCUITS is published monthly. Contributed papers may be of a tutorial or research nature, but the latter must be original and must not duplicate descriptions or derivations available elsewhere. This JOURNAL publishes papers of broad interest in the area of solid-state circuits. It provides device coverage in the areas of central importance to circuits and systems coverage in areas contiguous with circuit considerations. The subject matter should therefore relate to the analysis, design, and performance of solid-state circuits that may contain combinations of transistors, diodes, bulk-effect devices, magnetic devices, etc. The circuits may be digital, analog, microwave, or optoelectronic in nature; integrated circuits and large-scale integration are of principal interest.

Submission of a manuscript manifests the fact that it has received proper clearance from the author's company or institution, that it has not been copyrighted, published, submitted, or accepted for publication elsewhere, and that it will not be submitted elsewhere while under consideration. The manuscript text should not contain any commercial references, such as company names, university names, trademarks, commercial acronyms, or part numbers. All material not accepted will be returned.

Length

Authors should document their work in relation to the open literature. The following limits on length will be enforced:

1) Regular papers, up to 15 double-spaced typewritten pages in length, plus up to 10 pages, $8\frac{1}{2}''$ by $11''$, of figures.
2) Correspondence items of up to 3 double-spaced typewritten pages, plus not more than 2 pages of figures.

Styles for Manuscript

1) The manuscript should be typewritten using double or $1\frac{1}{2}$ space; use one side of the sheet only, single-column format; use 12 point font, 1 inch margins, and no more than 25 lines of text per page. Office-duplicated copies are acceptable.
2) Provide a carefully worded abstract of 100 to 200 words for papers and less than 50 words for correspondence. Name, address, telephone number, fax, and e-mail address of author(s) should appear with the abstract.
3) The style for organization of a paper, abbreviations, etc., can be determined from previous issues of this JOURNAL. Additional information is available on the web at http://www.ieee.org/organizations/pubs/transactions/information.htm.
4) References should appear in a separate bibliography at the end of the paper. References should be complete and in IEEE style.
   Style for papers: Author (with initials first), title, journal title, volume number, inclusive page numbers, month, year.
   Style for books: Author, title, location, publisher, year, page or chapter number (if desired). See this issue for further examples.
5) Figure captions should be on a separate sheet in proper style for typesetting.
6) Departures from the above style may delay publication.

Style for Illustrations

1) Originals of drawings and glossy print photographs should be sharp and of good contrast. Line drawings should be in black ink on a white background. Use $8\frac{1}{2} \times 11$ inch size sheets to simplify handling of the manuscript. Template lettering is recommended; typing on figures is not acceptable. Lettering should be large enough to permit legible reduction of the figure to column width, perhaps as much as 4:1.
2) Identify each illustration on the back or at the bottom of the front with the figure number and name of author(s). Indicate the top of a photograph. *Captions lettered on figures will be blocked out in reproduction in favor of typeset captions.*

Copyright

It is the policy of the IEEE to own the copyright to the technical contributions it publishes on behalf of the interests of the IEEE, its authors, and their employers, and to facilitate the appropriate reuse of this material by others. To comply with the U.S. Copyright Law, authors are required to sign an IEEE copyright transfer form (http://www.ieee.org/about/documentation/copyright) before publication. This form, a copy of which appears in the January issue of this JOURNAL, returns to authors and their employers full rights to reuse their material for their own purposes. Authors must submit a signed copy of this form with their manuscripts.

Submissions

There are two ways to submit a manuscript. The first is to send five paper copies of your manuscript to the Editor. Each copy should be complete with illustrations and should be accompanied by a separate sheet containing the address to which proofs and other correspondence can be sent. The second submission method is to send your manuscript in Portable Document Format (PDF) as an e-mail attachment to jssc-subm@ieee.org. In the body of the e-mail message, include the name, address, phone number, fax number, and e-mail address of the contact author, as well as the manuscript title and authors' names. Also be prepared to submit original illustrations immediately upon acceptance of your manuscript. In the case of regular papers, also be prepared to provide a brief technical biography and photograph of each author.

Review Process

The review process usually requires about three months. The author is then notified of the decision of the Editor or Associate Editor. The authors may be asked to modify the manuscript if it is not accepted or rejected in its original form. The elapsed time between receipt of a manuscript and publication is usually about 12 months.

Electronic Copy of Accepted Manuscripts

After a manuscript has been accepted for publication, a disk copy of the manuscript should be included along with the final paper copy of the manuscript. Please do not submit on a disk versions of manuscripts that have not been accepted for publication.

The following is a list of general guidelines for submission of a manuscript on a disk by prospective authors.

- The operating system and word-processing software used to produce your document should be noted on your disk and summarized on a final manuscript summary form (http://www.sscs.org/jssc/j_submit.pdf), which should be included with the disk. In the case of UNIX media, the method of extraction (i.e., tar, bar, dump) should also be noted.
- The hardcopy should match its companion disk exactly. Any changes that have been made to the hardcopy should be incorporated into your disk.
- The disk should be labeled with the file name(s) relating to the manuscript.
- No program files should be included on the disk.
- Package compact or floppy disks in such a way as to minimize possible damage in transit.
- Include a flat ASCII version on the disk with the word-processed version, if possible.
- Adhere to the accepted style of this JOURNAL as much as possible.

While the list of processing programs acceptable to IEEE electronic production is almost endless, there are preferred programs that, when used, enable the preservation of the greatest amount of information. TeX or LaTeX are preferred. An IEEE LaTeX style file, IEEEtran.sty, is obtainable by authors (http://www.ieee.org/organizations/pubs/authors.html) or e-mail a request to trans@ieee.org. The following points are important to remember when submitting disks in TeX or LaTeX. 1) Include all macros (\def) that are required to produce your document. 2) IEEE JOURNAL style dictates a 21-pica (3.5-in) column width. If mathematical phrases are produced with this in mind, they are apt to appear more aesthetically pleasing in the final version.

The IEEE will accept IBM-PC or Macintosh compatible compact disks (CDs), 3.5-in floppy disks, or zip disks. If in doubt, don't hesitate to inquire using trans@ieee.org.

Page Charges

After a manuscript has been accepted for publication, the author's company or institution will be requested to pay a charge of $110 per printed page to cover part of the cost of publication. Page charges for this IEEE JOURNAL, like those for journals of other professional societies, are not obligatory, nor is their payment a prerequisite for publication. The author will receive 100 free reprints without covers if the charge is honored. Detailed instructions will accompany the galley proof.

Digital Object Identifier 10.1109/JSSC.2005.844052

(Contents Continued from Front Cover)

BRIEF PAPERS

CORRESPONDENCE